

## THÈSE

présentée à

l'Université Paris XII - Val-de-Marne

pour l'obtention du grade de **Docteur en Sciences**

Spécialité : Chimie de la Pollution Atmosphérique et Physique de l'Environnement

par

**Isabelle PISON**

---

# Modélisation inverse pour l'optimisation des sources primaires de pollution atmosphérique à l'échelle régionale

---

Soutenue le 7 décembre 2005 devant le jury composé de :

Claire Granier	Rapporteur
Richard Ménard	Rapporteur
Gérard Ancellet	Président
Cécile Puel	Examineur
Gilles Bergametti	Examineur
Laurent Menut	Examineur

Thèse préparée au Laboratoire Inter-Universitaire des Systèmes Atmosphériques  
co-financée par le CNRS et Total

Mes remerciements vont à François Raulin, directeur du LISA ainsi qu'à Jean-Marie Flaud, qui lui a succédé : ils m'ont permis de travailler dans de très bonnes conditions pendant mes années de thèse.

Je remercie également le CNRS et Total, qui ont financé ce travail pendant trois ans, et en particulier Cécile Puel et Olivier Duclaux, avec qui les réunions d'avancement ont toujours été à la fois agréables et efficaces.

Je tiens à remercier les membres du jury, particulièrement Claire Granier et Richard Ménard, rapporteurs, et Gérard Ancellet, président, pour l'intérêt qu'ils ont porté à mon travail et leurs suggestions.

Je remercie Airparif, qui a fourni de nombreuses données indispensables à ce travail, ainsi que les personnes du KNMI, notamment Jan Fokke Meirink, qui m'ont accueillie pendant deux semaines pour des échanges très importants. Et évidemment, je suis redevable à Nadège Blond pour son travail et ses nombreuses explications : sans elle, mon travail n'aurait pu aboutir.

Bien sûr, mes remerciements les plus vifs vont à Laurent Menut et Gilles Bergammetti, qui m'ont encadrée tous les deux de façon très complémentaire au long de ce travail. Ils se sont relayés pour m'apporter leur soutien et leurs connaissances.

Enfin, je n'oublie pas tous ceux qui ont contribué à créer une ambiance de travail agréable : Nathalie, Benoît, Fanny et Jean-Michel, l'équipe modélisation au complet, Alma et ses collègues du LMD que j'ai brièvement rencontrés.

**«Dans un pays sous-développé, ne buvez jamais d'eau. Mais, dans un pays  
industrialisé, ne respirez jamais l'air.»**

Jonathan Raban cité par Claude Gagnière  
dans le Bouquin des citations (Robert Laffont, 2000)

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>16</b>
1.1	La pollution à l'échelle régionale . . . . .	16
1.1.1	Définition . . . . .	16
1.1.2	Historique . . . . .	17
1.1.3	Sources et puits . . . . .	19
1.1.3.1	Les émissions surfaciques et aériennes . . . . .	19
1.1.3.2	La chimie troposphérique . . . . .	20
1.1.3.3	La météorologie : les interactions physico-chimiques . . . . .	20
1.1.3.4	La turbulence . . . . .	21
1.1.3.5	Le dépôt . . . . .	21
1.1.4	La mesure de la pollution . . . . .	22
1.1.4.1	Moyens et objectifs . . . . .	22
1.1.4.2	Intérêt et difficultés . . . . .	23
1.1.5	Études, prévisions, scénarios et modèles numériques . . . . .	23
1.1.5.1	Nécessité de la modélisation numérique . . . . .	23
1.1.5.2	Ce qu'est un modèle déterministe . . . . .	24
1.1.5.3	Avantages et limites des modèles . . . . .	24
1.2	Les émissions surfaciques . . . . .	25
1.2.1	Émissions anthropiques et biogéniques . . . . .	25
1.2.1.1	Les émissions anthropiques . . . . .	25
1.2.1.2	Les émissions biogéniques . . . . .	25
1.2.2	La construction de cadastres «bottom-up» . . . . .	26
1.2.3	La mise en œuvre dans les modèles : les espèces-modèle . . . . .	27
1.3	Les objectifs de cette thèse . . . . .	27
1.4	Plan du manuscrit . . . . .	28
<b>2</b>	<b>Méthodes inverses</b>	<b>29</b>
2.1	État de l'art . . . . .	29
2.1.1	Principe général et historique . . . . .	29
2.1.2	Quelles sont les méthodes utilisées ? . . . . .	30
2.1.3	Quelles émissions ont été inversées ? . . . . .	33
2.2	Limitations des méthodes inverses et échelles d'application . . . . .	35
2.2.1	Limites dues aux mesures : toutes les informations nécessaires ne sont pas disponibles . . . . .	35
2.2.2	Limites dues au modèle : les erreurs dues au modèle n'ont pas une influence minimale . . . . .	37

2.2.3	Différence entre les échelles d'application . . . . .	39
2.3	Élaboration de notre méthode d'inversion . . . . .	40
2.3.1	Le vecteur d'état . . . . .	40
2.3.2	La matrice $\mathbf{R}$ . . . . .	40
2.3.3	La matrice $\mathbf{B}$ . . . . .	41
2.4	Conclusion . . . . .	41
<b>3</b>	<b>Cadre de modélisation</b>	<b>42</b>
3.1	Le modèle CHIMERE . . . . .	42
3.1.1	Principe général du modèle . . . . .	42
3.1.2	Les paramètres d'entrée . . . . .	43
3.1.2.1	La géométrie des domaines . . . . .	43
3.1.2.2	La météorologie . . . . .	43
3.1.2.3	Les conditions aux limites et les conditions initiales . . . . .	44
3.1.2.4	Les émissions . . . . .	44
3.1.3	Principes généraux des schémas et paramétrisations . . . . .	46
3.2	L'adjoint . . . . .	48
3.2.1	Principe . . . . .	48
3.2.2	Développement et validation . . . . .	48
3.3	L'optimiseur . . . . .	49
3.4	L'inverse . . . . .	49
3.4.1	Principe : imbrication des modules de calcul . . . . .	49
3.4.2	Le critère d'arrêt . . . . .	51
3.5	Un outil d'étude méthodologique : les cas académiques . . . . .	51
3.5.1	Principe . . . . .	51
3.5.2	Du simple au quasi-réel . . . . .	52
3.6	Validation de la gestion numérique . . . . .	53
3.6.1	Un modèle simple de cas académiques . . . . .	53
3.6.1.1	Maillage et météorologie . . . . .	53
3.6.1.2	Les émissions . . . . .	54
3.6.2	Résultats . . . . .	55
3.6.2.1	Cas formel 1 : précision théorique du système . . . . .	55
3.6.2.2	Cas formel 2 : un exemple simple de manque d'information . . . . .	56
3.7	Conclusion . . . . .	57
<b>4</b>	<b>Méthodologie d'optimisation avec krigeage</b>	<b>58</b>
4.1	Le krigeage . . . . .	58
4.1.1	Rappels théoriques . . . . .	58
4.1.2	Mise en œuvre . . . . .	59
4.2	Krigeage et inversion itérative . . . . .	59
4.3	Stratégie de validation : les cas académiques . . . . .	60
4.3.1	Principe . . . . .	60
4.3.2	Cas académiques élaborés . . . . .	60
4.3.3	Inversions . . . . .	62
4.4	Résultats . . . . .	62

4.4.1	Définition du score . . . . .	62
4.4.2	Nécessité du krigeage . . . . .	63
4.4.3	Impact du krigeage . . . . .	64
4.4.4	Lien entre contraintes et espèce à inverser . . . . .	66
4.4.4.1	Concentrations de NO . . . . .	66
4.4.4.2	Concentrations d’ozone . . . . .	67
4.5	Conclusion . . . . .	70
<b>5</b>	<b>Agrégation spatiale et autres paramètres : compléter la méthodologie</b>	<b>71</b>
5.1	L’agrégation spatiale . . . . .	71
5.1.1	Principe . . . . .	71
5.1.2	Mise en œuvre . . . . .	72
5.1.3	Validation par des cas académiques . . . . .	74
5.1.3.1	Cas académiques élaborés . . . . .	74
5.1.3.2	Inversions . . . . .	74
5.1.3.3	Résultats . . . . .	75
5.1.3.4	Conclusion . . . . .	77
5.2	Prise en compte de paramètres autres que les émissions . . . . .	78
5.2.1	Nécessité . . . . .	78
5.2.2	Mise en œuvre dans l’inverse : la matrice «de background» . . . . .	78
5.2.3	Mise en œuvre en amont de l’inverse . . . . .	81
5.2.3.1	Correction de biais . . . . .	81
5.2.3.2	Encadrement pour une incertitude connue . . . . .	82
5.3	Conclusion : la méthode complète . . . . .	82
5.3.1	L’inversion . . . . .	82
5.3.2	Stratégie générale . . . . .	83
<b>6</b>	<b>Application à des cas réels</b>	<b>85</b>
6.1	Le cadre de l’étude : l’Île-de-France . . . . .	85
6.2	Position du problème . . . . .	85
6.2.1	Impact des émissions 3D . . . . .	85
6.2.2	Comparaisons entre les mesures et le modèle . . . . .	86
6.2.2.1	Les données . . . . .	86
6.2.2.2	Les indicateurs statistiques . . . . .	89
6.2.2.3	Résultats statistiques . . . . .	90
6.3	Choix des cas à étudier . . . . .	97
6.3.1	Hypothèses . . . . .	97
6.3.1.1	Le modèle . . . . .	97
6.3.1.2	Le cadastre . . . . .	97
6.3.2	Le problème à inverser . . . . .	97
6.3.2.1	Émissions à optimiser . . . . .	97
6.3.2.2	Contraintes . . . . .	98
6.4	Le 7 août 1998 . . . . .	99
6.4.1	Choix de cette journée . . . . .	99
6.4.2	Impact du mélange vertical . . . . .	102
6.4.2.1	Coefficient de mélange turbulent . . . . .	102

6.4.2.2	Hauteur de couche-limite . . . . .	102
6.4.2.3	Conclusion . . . . .	104
6.4.3	Optimisation des émissions . . . . .	105
6.4.3.1	Inversion des flux de $\text{NO}_x$ . . . . .	105
6.4.3.2	Impact sur les concentrations d'ozone . . . . .	109
6.4.4	Conclusion . . . . .	112
6.5	Du 16 au 18 juillet 1999 . . . . .	113
6.5.1	Choix de cette période . . . . .	113
6.5.2	Inversion des émissions de $\text{NO}_x$ . . . . .	117
6.5.3	Conclusion . . . . .	123
6.6	Inversions climatologiques . . . . .	124
6.6.1	Choix de la période . . . . .	124
6.6.2	Inversion des émissions de $\text{NO}_x$ . . . . .	124
6.6.3	Synthèse des résultats . . . . .	138
6.7	Conclusion . . . . .	140
<b>7</b>	<b>Conclusions et perspectives</b>	<b>142</b>
7.1	Conclusions . . . . .	142
7.2	Perspectives . . . . .	145
<b>A</b>	<b>La photochimie dans la troposphère</b>	<b>148</b>
A.1	L'équilibre photochimique . . . . .	148
A.2	Le cycle des radicaux . . . . .	148
A.2.1	Production d'ozone . . . . .	149
A.2.2	Sources et puits pour les radicaux . . . . .	149
A.2.3	Les régimes chimiques . . . . .	150
<b>B</b>	<b>La campagne ESQUIF</b>	<b>151</b>
B.1	La campagne de mesure . . . . .	151
B.2	Caractéristiques des POI . . . . .	152
<b>C</b>	<b>L'adjoint</b>	<b>153</b>
C.1	Principe . . . . .	153
C.2	Programmation . . . . .	155
C.2.1	Principe d'écriture d'un code adjoint . . . . .	155
C.2.2	Gestion du temps dans l'adjoint . . . . .	157
C.2.3	Vérification informatique du code adjoint . . . . .	158
C.3	Schéma de transport de Van Leer : codes direct et adjoint . . . . .	159
C.3.1	Code direct . . . . .	159
C.3.2	Adjointisation . . . . .	159
C.3.3	Code adjoint . . . . .	161
<b>D</b>	<b>Inverse</b>	<b>162</b>

<b>E</b>	<b>Le krigage</b>	<b>164</b>
E.1	Le krigage . . . . .	164
E.1.1	Principe . . . . .	164
E.1.2	Les poids . . . . .	164
E.2	Le krigage des innovations . . . . .	165
E.2.1	Principe . . . . .	165
E.2.2	Les poids . . . . .	166
E.3	Krigage intrinsèque . . . . .	166
<b>F</b>	<b>Méthode itérative et optimalité</b>	<b>167</b>
F.1	Définitions générales . . . . .	167
F.2	Approche classique . . . . .	167
F.3	Développement : méthode itérative . . . . .	167
F.3.1	Krigage des concentrations . . . . .	167
F.3.2	Inversion des émissions . . . . .	168
F.3.3	Deuxième cycle krigage-optimisation . . . . .	168
F.3.4	N-ième cycle krigage-optimisation . . . . .	168
F.3.5	Convergence . . . . .	168
<b>G</b>		<b>169</b>



# Table des figures

1.1	<i>Ordres de grandeur de l'épaisseur des couches basses de l'atmosphère.</i>	17
1.2	<i>Temps de vie de quelques espèces chimiques et échelle spatiale correspondante (d'après L. Menut).</i>	18
1.3	<i>Schéma récapitulant les différents processus dynamiques et chimiques (d'après L. Menut).</i>	19
1.4	<i>Évolution temporelle de la couche-limite (d'après L. Menut).</i>	21
1.5	<i>Schéma de principe d'un modèle de chimie-transport (CTM) déterministe.</i>	24
2.1	<i>Principe de la modélisation inverse. CTM = Chemistry Transport Model</i>	32
3.1	<i>Cartes des domaines de simulation CHIMERE</i>	43
3.2	<i>Différence entre la méthode de séparation des opérateurs («operator splitting») et la méthode utilisée dans CHIMERE («Prod-Loss budget»).</i>	47
3.3	<i>Principe de l'adjoint d'un modèle de chimie-transport.</i>	48
3.4	<i>Principe de l'inversion d'une simulation.</i>	50
3.5	<i>Principe d'élaboration d'un cas académique</i>	52
3.6	<i>Évolution de la température à 2 mètres sur 24 heures.</i>	54
3.7	<i>Niveaux verticaux et évolution de la hauteur de couche-limite en 24 heures.</i>	54
3.8	<i>Concentrations surfaciques de NO et écarts (ppb) entre concentrations a priori et «réelles» à 10 heures.</i>	55
3.9	<i>Coefficients correctifs pour les sources 1 et 2 à chaque heure de la fenêtre temporelle pour le cas formel 1.</i>	56
3.10	<i>Coefficients correctifs pour les sources 1 et 2 à chaque heure de la fenêtre temporelle pour le cas formel 2.</i>	56
4.1	<i>Exemple de résultats obtenus en Île-de-France avec le réseau AIRPARIF, le modèle direct CHIMERE et la méthode de krigeage INK.</i>	60
4.2	<i>Principe d'élaboration d'un cas académique avec krigeage.</i>	61
4.3	<i>Cadastre d'émission «réel» : flux de NO à 9 heures le 7 août 1998. <math>x</math> = flux perturbé pour créer le cadastre a priori.</i>	62
4.4	<i>Écart entre le flux optimisé et le flux réel (en % par rapport au flux réel) : comparaison des résultats obtenus en utilisant trois types de contraintes différents. Écart entre les flux a priori et les flux réels = -50%.</i>	65
4.5	<i>Scores moyens (%) après un à trois cycles krigeage-optimisation.</i>	66
4.6	<i>Évolution des concentrations de NO «réelles», a priori et optimisées pour 0100 avec les contraintes obtenues à partir de mesures à <math>\pm=10\%</math>, dans une maille située au-dessus d'un des flux inversés.</i>	67

4.7	Écart (ppb) entre les concentrations simulées avec le cadastre a priori o100 et les concentrations «réelles» à l'heure du pic. . . . .	68
4.8	Évolution des concentrations d'ozone «réelles», a priori et optimisées pour o100 avec les contraintes obtenues à partir de mesures à $\pm=15\%$ , au cœur du panache. . . . .	68
4.9	Cadastre d'émission «réel» : flux total de COV à 9 heures le 7 août 1998. $x$ = flux perturbé pour créer le cadastre a priori. . . . .	69
4.10	Écart (ppb) entre les concentrations d'ozone simulées avec le cadastre a priori o100 pour les COV et les concentrations «réelles» à l'heure du pic. . . . .	69
5.1	Schéma d'un découpage en grandes régions selon la direction du vent. . .	73
5.2	Zones obtenues pour le cas o50-58 à 8 heures (29 flux), selon la tolérance $\tau_E$ choisie. . . . .	75
5.3	Différence (en %) avec le cadastre «réel» de NO pour le cas o-58 à 8 heures. . . . .	77
5.4	Concentrations surfaciques de NO et écarts (ppb) entre concentrations a priori et «réelles» à 10 heures. . . . .	80
5.5	Coefficients correctifs pour les émissions (pour la source 2) et les températures à chaque heure de la fenêtre temporelle quand les flux d'émissions sont indépendants. . . . .	80
5.6	Coefficients correctifs pour les émissions et les températures à chaque heure de la fenêtre temporelle quand les flux d'émissions sont regroupés. . . . .	81
5.7	Tous les éléments de notre méthode d'inversion. . . . .	83
6.1	Cartes composites d'écarts (ppb) entre les concentrations d'ozone de la simulation prenant en compte les émissions des avions jusqu'à 3000 mètres et de la simulation de référence sans émission tridimensionnelle : pour chaque maille, l'écart maximal (négatif et positif) obtenu sur toute la durée de la simulation de la POI 2 à tous les niveaux est représenté. L'échelle est choisie de telle sorte qu'on ne prend pas en compte les 10% les plus faibles. CdG = aéroport Charles-de-Gaulle. . . . .	87
6.2	Stations de mesures AIRPARIF. Le type de station est indiqué en couleur : rouge = urbain, bleu = observation, vert = rural, jaune = trafic. Les mesures effectuées sont indiquées par le symbole : croix = $O_3$ , NO et $NO_2$ , cercle = $O_3$ seulement, triangle = NO et $NO_2$ seulement. . . . .	88
6.3	Répartition géographique de la corrélation calculée pendant l'été 1999 pour les concentrations d'ozone (ppb). . . . .	91
6.4	Évolution du biais selon l'heure de la journée pendant les étés 1998 et 1999 pour les concentrations d'ozone (ppb). En été, TU = heure légale - 2 heures. . . . .	92
6.5	Évolution de la corrélation selon l'heure de la journée pendant les étés 1998 et 1999 pour les concentrations d'ozone (ppb). En été, TU = heure légale - 2 heures. . . . .	93
6.6	Répartition géographique de la corrélation calculée sur tout l'été en 1998 et 1999 pour les concentrations de NO (ppb). . . . .	94
6.7	Évolution du biais selon l'heure de la journée pendant l'été 1999 pour les concentrations de NO (ppb). En été, TU = heure légale - 2 heures. . . . .	96

6.8	<i>Cadastre d'émissions a priori : flux de NO le 7 août 1998 à 8 heures TU. En été, TU = heure légale - 2 heures. . . . .</i>	98
6.9	<i>Concentrations d'O<sub>3</sub> (en ppb) en surface le 7 août à 15 heures TU. Les flèches rouges indiquent le sens et l'intensité du vent, les étiquettes correspondent aux valeurs relevées par des stations de mesure AIRPARIF. En été, TU = heure légale - 2 heures. . . . .</i>	99
6.10	<i>Comparaison entre les concentrations de monoxyde d'azote mesurées par le réseau AIRPRIF et calculées par le modèle pour le 7 août 1998. En été, TU = heure légale - 2 heures. . . . .</i>	100
6.11	<i>Comparaison entre les concentrations d'ozone mesurées par des stations AIRPARIF et calculées par le modèle le 7 août 1998. La Zone Rurale Nord-Est correspond à la station de Montgé-en-Goële (au vent) et la Zone Rurale Sud-Ouest, à la station de la forêt de Rambouillet (sous le vent) - voir Figure 6.2 page 88 pour l'emplacement des stations. En été, TU = heure légale - 2 heures. . . . .</i>	101
6.12	<i>Concentrations de monoxyde d'azote et d'ozone (ppb) obtenues avec différentes valeurs du coefficient de mélange turbulent <math>K_z</math>. En été, TU = heure légale - 2 heures. . . . .</i>	103
6.13	<i>Concentrations de monoxyde d'azote et d'ozone (ppb) obtenues avec différentes HCL et valeurs de <math>K_z</math>. En été, TU = heure légale - 2 heures. . . .</i>	104
6.14	<i>Profil temporel moyen sur tout le domaine du flux d'émission de NO le 7 août 1998. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées. . . . .</i>	105
6.15	<i>Répartition spatiale des corrections pour les NO<sub>x</sub> et intensité des émissions de NO (représentant 90% des flux de NO<sub>x</sub>) le 7 août 1998. En été, TU = heure légale - 2 heures. . . . .</i>	106
6.16	<i>Évolutions temporelles des émissions et des concentrations de monoxyde d'azote le 7 août 1998 dans la maille Paris Sud-Ouest. Les concentrations d'une heure donnée résultent des émissions de l'heure précédente. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées. . . . .</i>	107
6.17	<i>Évolutions temporelles des émissions et des concentrations de monoxyde d'azote le 7 août 1998 dans la maille de Tremblay. Les concentrations d'une heure donnée résultent des émissions de l'heure précédente. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées. . . . .</i>	108
6.18	<i>Différence (ppb) entre les concentrations d'ozone optimisées et a priori le 7 août 1998 à 15 heures TU. En été, TU = heure légale - 2 heures. . . . .</i>	110
6.19	<i>Profil temporel moyen sur tout le domaine du flux total d'émission de COV le 7 août 1998. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées. . . . .</i>	111
6.20	<i>Différence (ppb) entre les concentrations d'ozone optimisées et a priori le 7 août 1998 à 15 heures TU. En été, TU = heure légale - 2 heures. . . . .</i>	111

6.21	Concentrations d'O <sub>3</sub> (en ppb) en surface les 16, 17 et 18 juillet 1999 à 15 heures TU. Les flèches noires indiquent le sens et l'intensité du vent, les étiquettes correspondent aux valeurs relevées par des stations de mesure AIRPARIF. En été, TU = heure légale - 2 heures. . . . .	114
6.22	Comparaison entre les concentrations de monoxyde d'azote mesurées par des stations AIRPARIF et calculées par le modèle du 16 au 18 juillet 1999. En été, TU = heure légale - 2 heures. . . . .	115
6.23	Comparaison entre les concentrations d'ozone mesurées par des stations AIRPARIF et calculées par le modèle du 16 au 18 juillet 1999. La Zone Rurale Nord-Ouest correspond à la station de Frémainville et la Zone Rurale Est, à la station de Saints - voir Figure 6.2 page 88 pour l'emplacement des stations. En été, TU = heure légale - 2 heures. . . . .	116
6.24	Hauteur de couche-limite pendant la POI 6 d'ESQUIF : comparaison entre les mesures et les valeurs fournies au modèle par l'ECMWF. En été, TU = heure légale - 2 heures. . . . .	117
6.25	Profil temporel moyen sur tout le domaine du flux d'émission de NO du 16 au 18 juillet 1999. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées. . . . .	118
6.26	Répartition spatiale des corrections pour les NO <sub>x</sub> et intensité des émissions de NO (représentant 90% des flux de NO <sub>x</sub> ) les 16, 17 et 18 juillet 1999. En été, TU = heure légale - 2 heures. . . . .	119
6.27	Évolutions temporelles des émissions et des concentrations de monoxyde d'azote du 16 au 18 juillet 1999 dans la maille Paris Sud-Ouest. Les concentrations d'une heure donnée résultent des émissions de l'heure précédente. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées. . . . .	121
6.28	Évolutions temporelles des émissions et des concentrations de monoxyde d'azote du 16 au 18 juillet 1999 dans la maille de Tremblay. Les concentrations d'une heure donnée résultent des émissions de l'heure précédente. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées. . . . .	122
6.29	Illustrations de l'impact des conditions aux limites sur l'information concernant les émissions locales contenue dans l'écart ( $\Delta_{fg}$ ) entre concentrations mesurées et simulées. . . . .	129
6.30	Différence (en %) entre les profils temporels moyens sur tout le domaine des flux d'émission de NO optimisés avec la hauteur de couche-limite fournie par ECMWF et avec la hauteur de couche-limite corrigée grâce aux mesures. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées. . . . .	130
6.31	Profils temporels a priori des flux d'émission de NO moyennés sur tout le domaine, selon les différents types de jour pour les mois de juillet et août 1998 et 1999. En été, TU = heure légale - 2 heures. . . . .	131
6.32	Profil temporel médian des flux d'émission de NO moyennés sur tout le domaine, obtenu sur les périodes du 1er juillet au 31 août 1998 et 1999. En été, TU = heure légale - 2 heures. . . . .	131

6.33	<i>Distribution en fréquence (pour des classes de 0,05) des coefficients de correction obtenus à chaque heure inversée sur la période du 1er juillet au 31 août 1998 et 1999. Pour plus de lisibilité, la courbe indiquant la distribution de chaque heure est décalée vers le haut d'une unité de plus que la précédente.</i>	133
6.34	<i>Représentation spatiale des corrections les plus fréquemment appliquées aux flux d'émission de <math>\text{NO}_x</math> sur les périodes du 1er juillet au 31 août 1998 et 1999. Classes de 0,05 non représentées individuellement. En été, TU = heure légale - 2 heures.</i>	134
6.35	<i>Profils temporels médians des flux d'émission de NO optimisés moyennés sur tout le domaine, obtenu sur les mois de juillet et août 1998 et 1999. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.</i>	135
6.36	<i>Profil temporel médian sur tout le domaine des flux d'émission de NO pour le type «semaine» obtenu sur les périodes du 1er juillet au 31 août 1998 et 1999. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.</i>	135
6.37	<i>Profils temporels médians sur tout le domaine des flux d'émission de NO pour les types «samedi» et «dimanche ou jour férié» obtenus sur les périodes du 1er juillet au 31 août 1998 et 1999. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.</i>	137
6.38	<i>Profils temporels médian sur tout le domaine des flux d'émission de NO optimisés pour les jours de type «semaine» obtenus sur les mois de juillet et août 1998 et 1999. En été, TU = heure légale - 2 heures.</i>	139
C.1	<i>Schéma de principe de la gestion du temps dans l'adjoint</i>	157
C.2	<i>Schéma de principe de la vérification du codage du modèle adjoint</i>	158
D.1	<i>Algorithme de l'inverse de CHIMERE.</i>	163

# Liste des tableaux

2.1	<i>Études en modélisation inverse des émissions à l'échelle régionale. . . . .</i>	34
4.1	<i>Scores moyens obtenus pour l'optimisation des émissions de NO<sub>x</sub> avec des concentrations d'ozone ou de NO comme contraintes. x = convergence non atteinte. . . . .</i>	63
4.2	<i>Scores moyens (avec leur variance) obtenus pour l'optimisation des émissions de NO<sub>x</sub> avec des mesures ou des analyses de NO comme contraintes. . . . .</i>	64
5.1	<i>Flux perturbés et différence en % entre les cadastres a priori et «réel» entre 8 et 9 heures. Le nombre de flux perturbés est le total sur les deux heures. . . . .</i>	74
5.2	<i>Scores moyens obtenus pour l'optimisation des émissions de NO<sub>x</sub> avec des concentrations de NO comme contraintes, selon la dimension du problème (cas o50-58). x = convergence non atteinte. . . . .</i>	75
5.3	<i>Dimension permettant d'obtenir la convergence pour l'optimisation des émissions de NO<sub>x</sub> avec des concentrations de NO comme contraintes, selon le nombre de flux du problème. . . . .</i>	76
6.1	<i>Valeurs des variables statistiques (biais, moyenne quadratique (rms) et corrélation (cor)) calculées sur tout l'été en 1998 et 1999 pour les concentrations d'ozone (en ppb) dans les différentes mailles où se trouvent des stations de mesure. . . . .</i>	90
6.2	<i>Valeurs des variables statistiques (biais, moyenne quadratique (rms) et corrélation (cor)) calculées sur tout l'été en 1998 et 1999 pour les concentrations de NO (en ppb) dans les différentes mailles où se trouvent des stations de mesure. . . . .</i>	95
6.3	<i>Hauteur de couche-limite (m) mesurée et utilisée par le modèle le 7 août 1998. . . . .</i>	102
6.4	<i>Valeurs des variables statistiques (biais, moyenne quadratique (rms) et corrélation (cor)) calculées sur les mois de juillet et août en 1998 et 1999 pour les concentrations de NO (en ppb) a priori dans les différentes mailles où se trouvent des stations de mesure (x = pas de mesure à cette période) et écarts (en % de la valeur a priori) entre les valeurs calculées pour les concentrations de NO a priori et optimisées. <math>\Delta X = \frac{X_{\text{a priori}} - X_{\text{optimisé}}}{X_{\text{a priori}}} \times 100</math> . . . . .</i>	126

6.5	<i>Valeurs des variables statistiques (biais, moyenne quadratique (rms) et corrélation (cor)) calculées sur les mois de juillet et août en 1998 et 1999 pour les concentrations d’ozone (en ppb) a priori dans les différentes mailles où se trouvent des stations de mesure (x = pas de mesure à cette période) et écarts (en % de la valeur a priori) entre les valeurs calculées pour les concentrations d’ozone a priori et optimisées.</i>	
	$\Delta X = \frac{X_{\text{apriori}} - X_{\text{optimisé}}}{X_{\text{apriori}}} \times 100$	127
6.6	<i>Conséquence sur les émissions optimisées des différentes combinaisons d’erreurs sur le cadastre a priori et la hauteur de couche-limite.</i>	128
B.1	<i>Récapitulatif des douze POI d’ESQUIF. La colonne «origine de la masse d’air» indique l’origine continentale ou maritime de celle-ci. Les «re-marques» mettent en relief la spécificité de chacune des situations météorologiques.</i>	152

# Chapitre 1

## Introduction

### 1.1 La pollution à l'échelle régionale

#### 1.1.1 Définition

D'après la Loi sur l'Air et l'Utilisation Rationnelle de l'Énergie du 30 décembre 1996 (LAURE, loi numéro 96-1236), «*constitue une **pollution atmosphérique** [...] l'introduction par l'homme, directement ou indirectement, dans l'atmosphère et les espaces clos, de substances ayant des conséquences préjudiciables de nature à mettre en danger la santé humaine, à nuire aux ressources biologiques et aux écosystèmes, à influencer sur les changements climatiques, à détériorer les biens matériels, à provoquer des nuisances olfactives excessives*» (souligné par nous, disponible sur [www.legifrance.gouv.fr](http://www.legifrance.gouv.fr)). Toutes les substances qualifiées de polluants atmosphériques ne sont pas étrangères à l'atmosphère et elles peuvent aussi avoir des conséquences positives pour la vie, comme l'ozone présent dans la stratosphère ou le dioxyde de carbone. Une substance présente dans l'atmosphère devient donc un polluant si sa concentration est modifiée de telle sorte que «des conséquences préjudiciables» apparaissent.

La pollution atmosphérique résulte, d'après la loi, de «l'introduction par l'homme, directement ou indirectement, dans l'atmosphère» de polluants. Cependant, les polluants ne sont pas nécessairement différents des substances émises par des phénomènes naturels. Si les CFC (chlorofluorocarbones) sont produits et émis uniquement par l'homme, des activités humaines comme l'agriculture ou l'élevage conduisent à l'émission de méthane ou de poussières qui sont aussi naturellement émises par les bactéries des marais ou les déserts.

La définition donnée par la LAURE prend en compte les différentes échelles spatio-temporelles auxquelles la pollution atmosphérique peut survenir : (i) la macro-échelle, qui est celle des phénomènes globaux comme les changements climatiques, (ii) la méso-échelle, celle des phénomènes qui concernent un continent ou une région comme les pluies acides qui nuisent aux ressources biologiques et aux écosystèmes, (iii) la micro-échelle, qui est celle des phénomènes locaux comme la détérioration des biens matériels, par exemple les monuments, ou les nuisances olfactives excessives, notamment sous le vent de certains types d'usines.

Cette thèse a pour cadre la pollution atmosphérique **à l'échelle régionale** :

– temporellement, les espèces chimiques d'intérêt en pollution régionale ont des temps



de vie de l'ordre de la fraction de seconde à quelques jours, comme indiqué sur la Figure 1.2 qui met en relation échelles spatiale et temporelle ; on remarquera notamment l'ozone ( $O_3$ ) et les oxydes d'azote ( $NO_x$ ) qui jouent un rôle prépondérant dans la pollution dite photo-oxydante décrite page 20. Les polluants considérés ici quantitativement sont des **gaz-trace** : leurs concentrations sont au maximum de l'ordre du ppm (partie par million).

- verticalement, les processus physico-chimiques qui intéressent ces espèces ont lieu dans la couche limite atmosphérique (CLA) et la partie de la troposphère libre située juste au-dessus. La CLA est la couche atmosphérique sous l'influence des processus de surface ; c'est donc la partie basse de l'atmosphère dont l'écoulement est influencé par les effets thermiques et dynamiques dus à la surface. Son extension verticale dépend des conditions météorologiques (vent, température) mais reste comprise entre 100 et 3000 mètres maximum. Au-dessus de la CLA, se trouvent la troposphère libre, puis la stratosphère (dès 10 km d'altitude) (Figure 1.1 d'après Stull (1988)).
- horizontalement, leurs temps de vie permettent à certaines espèces d'être transportées sur plusieurs dizaines de kilomètres. L'intérêt portera donc sur une grande agglomération et son proche environnement soit un domaine de type européen de l'ordre de la centaine de kilomètres de côté.

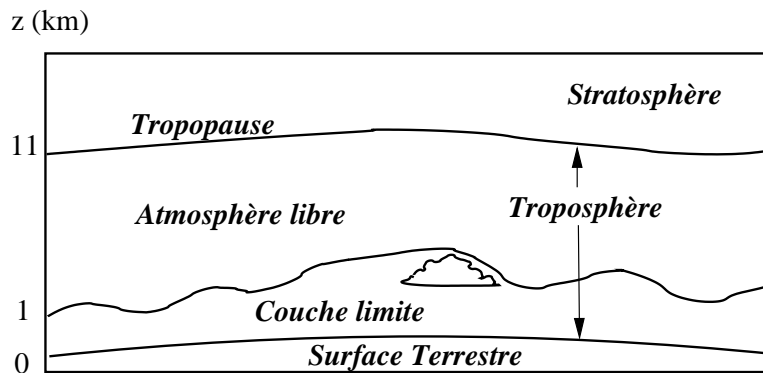


FIG. 1.1 – Ordres de grandeur de l'épaisseur des couches basses de l'atmosphère.

### 1.1.2 Historique

Au XX<sup>e</sup> siècle, la prise de conscience des problèmes posés par la pollution atmosphérique à l'échelle régionale a été ponctuée d'épisodes dramatiques pour la population ou son environnement.

En décembre 1930, dans la vallée de la Meuse près de Liège, en octobre 1948 à Donora, en Pennsylvanie, en décembre 1952 à Londres, des pics de morbidité et de mortalité ont été observés. Dans tous les cas, une situation météorologique anticyclonique accompagnée de brouillard s'est maintenue plusieurs jours, ce qui a permis l'accumulation des polluants émis par les industries et la ville. Le dioxyde de soufre, en présence du brouillard, s'est transformé en acide sulfurique, particulièrement agressif. Le terme de «smog», forgé au début du siècle pour décrire l'atmosphère londonienne faite de «smoke» (fumée) et de «fog» (brouillard), a été utilisé pour désigner ce type d'épisodes de pollution.

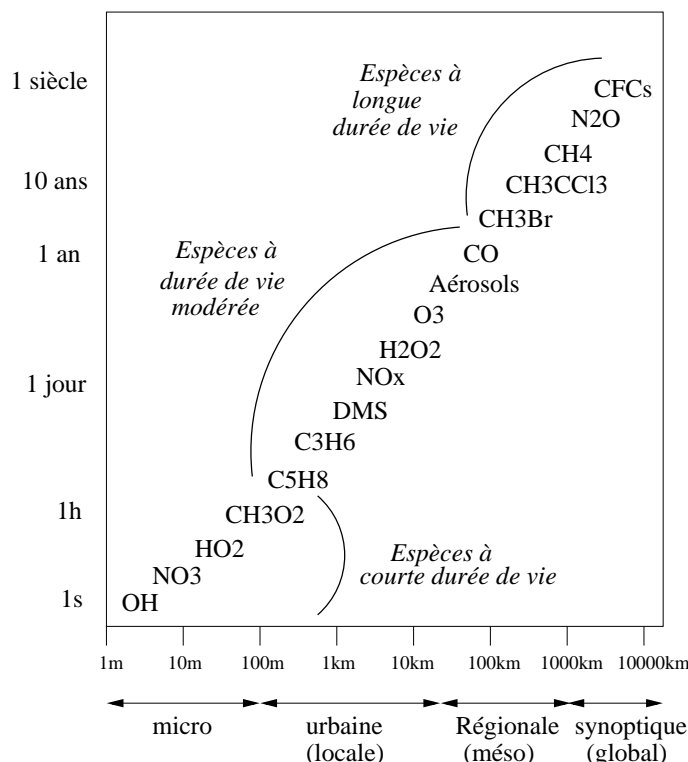


FIG. 1.2 – Temps de vie de quelques espèces chimiques et échelle spatiale correspondant (d'après L. Menut).

Dans le bassin de Los Angeles, dès les années 50, une baisse de la visibilité, des irritations des appareils oculaire et respiratoire humains et des effets néfastes sur les végétaux ont été relevés. La forte densité de population, la forte circulation automobile et l'industrialisation y ont en effet rencontré une situation géographique et météorologique favorable à la pollution dite «photo-oxydante».

En France, la prise de conscience des problèmes de pollution atmosphérique par le biais des pluies acides a amené, dans les années 80, aux premières mesures de réduction des émissions (de dioxyde de soufre puis d'autres polluants) et parallèlement a conduit au développement des activités de surveillance et de recherche. L'existence de phénomènes tels que le développement de panaches de pollution sous le vent des grandes agglomérations a commencé à être soupçonnée. Enfin, en 1996, la LAURE a rendu obligatoire la surveillance de la qualité de l'air au niveau local, sur l'ensemble du territoire depuis l'an 2000, ainsi que l'information du public. Des objectifs de qualité de l'air et des seuils variés (seuils d'information ou d'alerte de la population notamment) ont été établis. Une procédure d'alerte a été instaurée ; elle est gérée par le Préfet, qui doit informer le public et peut prendre des mesures d'urgence telles que la restriction de la circulation ou de certaines activités industrielles en cas de dépassement de seuil. L'élaboration d'un Plan Régional de la Qualité de l'Air (PRQA), de Plans de Protection de l'Atmosphère (PPA) et d'un Plan de Déplacement Urbain (PDU) pour toute agglomération de plus de 100 000 habitants est également prévue par la loi.

La perception et la connaissance de la pollution atmosphérique ont donc évolué rapide-

ment et la pollution elle-même a changé avec les années. Comme le montrent les cas dramatiques décrits plus haut, le dioxyde de soufre émis par les industries était un des polluants qui posaient le plus de problèmes jusque dans les années quatre-vingts. Les émissions de plomb, présent dans les carburants jusqu'en 2000, étaient également particulièrement préoccupantes. Depuis quelques années, sont devenus inquiétants les pics d'ozone en été et de  $\text{NO}_x$  en hiver. L'ozone est un des plus forts oxydants connus et il a un impact négatif sur l'environnement et la santé. Le terme de  $\text{NO}_x$  désigne la «famille» regroupant NO et  $\text{NO}_2$ . Ce dernier est un gaz irritant pour les voies pulmonaires et peut par exemple provoquer des crises chez les asthmatiques. Enfin, de nouvelles pistes d'étude se dessinent depuis peu : les particules (ou PM) commencent à être mesurées et modélisées ; des questions surgissent au sujet des niveaux de fond (concentration moyenne représentative d'un secteur géographique entier, ambiance de pollution générale) de l'ozone dans la troposphère dont la valeur, de 15 à 20 ppb (partie par milliard) à la fin du XIX<sup>e</sup> siècle, a aujourd'hui doublé dans l'hémisphère Nord (Vingarzan 2004).

### 1.1.3 Sources et puits

Les espèces présentes dans l'atmosphère subissent des transformations liées à des processus dynamiques et chimiques. Ces processus peuvent être séparés en deux catégories selon leur rôle : les sources et les puits. Les émissions constituent une source, le dépôt, un puits et la situation météorologique (dynamique), la diffusion turbulente et la chimie sont à la fois source et puits.

Pour comprendre l'évolution des concentrations des différentes espèces chimiques, il est donc nécessaire de comprendre leurs sources et puits (Figure 1.3).

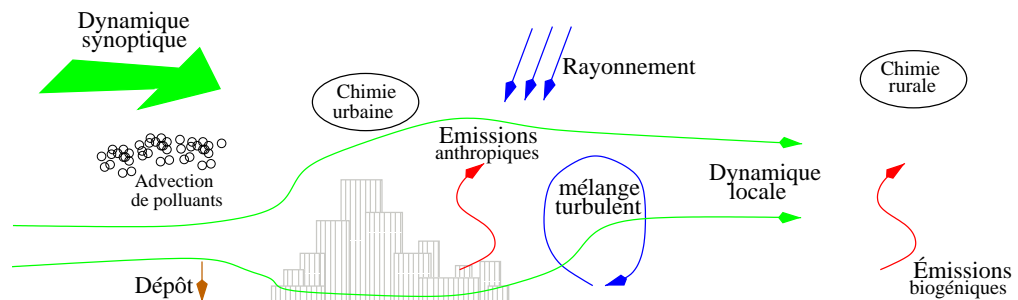


FIG. 1.3 – Schéma récapitulant les différents processus dynamiques et chimiques (d'après L. Menut).

#### 1.1.3.1 Les émissions surfaciques et aériennes

Les **émissions** constituent l'unique source de **polluants primaires** c'est-à-dire les polluants directement injectés dans l'atmosphère. La majeure partie des émissions est surfacique c'est-à-dire que les espèces sont émises à moins d'une dizaine de mètres du sol. Cependant, les cheminées d'usines peuvent atteindre plusieurs dizaines de mètres de hauteur et les avions traversent toute la CLA. Les émissions, qui constituent l'enjeu de ce travail, sont décrites de façon plus détaillée page 25.

### 1.1.3.2 La chimie troposphérique

Les espèces émises dans l'atmosphère peuvent réagir entre elles et mener à la production de polluants dits secondaires. On parle de **pollution photo-oxydante** lorsque les réactions chimiques en phase gazeuse aboutissant à la production de polluants secondaires (qui augmentent le pouvoir oxydant de l'atmosphère) ont été initiées par l'énergie solaire.

L'ensemble des réactions chimiques qui relient des émissions de polluants primaires aux concentrations de polluants secondaires forme un système non-linéaire faisant intervenir un grand nombre d'espèces aux temps de vie couvrant plus de dix ordres de grandeur (voir Figure 1.2 pour quelques exemples).

Pour l'étude de la pollution photo-oxydante (principalement estivale sous les latitudes tempérées), l'**ozone**, les **oxydes d'azote (NO et NO<sub>2</sub>, notés NO<sub>x</sub>)** et les **composés organiques volatils (COV)** sont particulièrement intéressants. Les NO<sub>x</sub> et les COV font partie des précurseurs de l'ozone : après émission, ils participent aux réactions de production et destruction de l'ozone, polluant secondaire produit par photolyse.

Les réactions principales du cycle de l'ozone sont reproduites en Annexe A. Près des sources émettrices, la masse totale de NO<sub>x</sub> est principalement constituée de NO ; lorsque l'on s'éloigne des sources, le NO se convertit peu à peu en NO<sub>2</sub> (voir réactions chimiques en Annexe A). C'est la concentration de NO qui détermine les réactions dominantes. La photochimie d'un environnement pollué c'est-à-dire riche en NO (plus de 10 à 30 ppb d'après Atkinson (2000)), typiquement en milieu urbain, diffère de celle d'un environnement non pollué (pauvre en NO), typiquement en milieu rural (Kleinman 1997).

En environnement non pollué, la faible quantité de NO disponible limite la formation d'ozone. Une augmentation de la concentration de NO, due à l'augmentation des émissions par exemple, contribue donc directement à la production d'ozone. Au contraire, en environnement pollué, où le système chimique est saturé en NO, une augmentation des émissions de NO<sub>x</sub>, à concentrations de COV égales, réduit la formation d'ozone. Si les émissions de COV augmentent en proportion des émissions de NO<sub>x</sub>, la production d'ozone continue. Celle-ci est donc sensible au **rapport COV/NO<sub>x</sub> des émissions** plutôt qu'aux valeurs des concentrations (Honoré et al. 2000; Honoré et Vautard 2000). On définit différents **régimes chimiques** selon qu'une augmentation des émissions de NO<sub>x</sub> (à émissions de COV constantes) conduit à une augmentation de la production d'ozone (régime NO<sub>x</sub> limité) ou à sa diminution (régime COV limité) (Sillman 1999).

### 1.1.3.3 La météorologie : les interactions physico-chimiques

Le terme d'advection désigne les mouvements horizontaux de l'atmosphère, liés le plus souvent au vent ; le terme de convection est utilisé pour les mouvements verticaux, en partie dus aux échanges de chaleur.

Le vent peut amener en un point des polluants émis ou produits bien plus loin. À l'échelle régionale, l'advection de masses d'air polluées s'ajoutant à la production locale peut déterminer de très intenses épisodes de pollution à l'ozone par exemple. De même, la dispersion des polluants émis ou produits sur place dépend de la vitesse et de la direction du vent. Un vent faible, une situation de recirculation peuvent amener à des concentrations particulièrement élevées.

La hauteur de la CLA (définie page 17) est déterminée par les profils verticaux de température et de vent. Elle délimite le volume dans lequel les émissions surfaciques sont di-

luées et les polluants mélangés. Son évolution au cours de la journée a donc une influence directe sur les concentrations (Figure 1.4). Lorsque la CLA est instable, la convection permet un mélange des polluants quasiment homogène dans toute la hauteur de la couche.

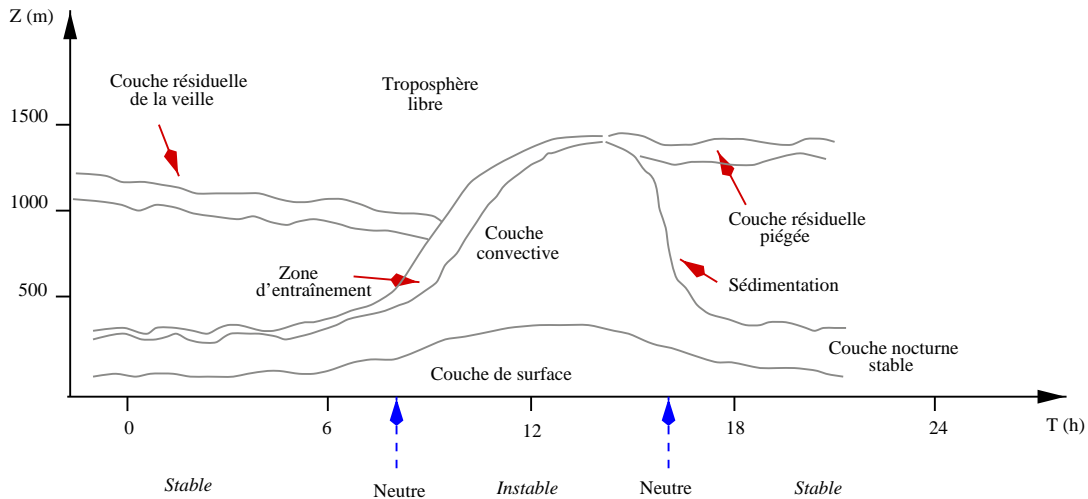


FIG. 1.4 – Évolution temporelle de la couche-limite (d'après L. Menut).

L'intensité du rayonnement solaire parvenant dans la CLA dépend de la quantité de nuages et de leur nature. Elle détermine la quantité d'énergie lumineuse disponible pour les réactions photolytiques, comme celles qui interviennent au cours du cycle de production/destruction de l'ozone.

La situation météorologique est donc déterminante dans le développement d'une situation de pollution atmosphérique à l'échelle régionale.

#### 1.1.3.4 La turbulence

À des échelles spatio-temporelles inférieures à la dizaine de mètres et à la minute, les mouvements de l'atmosphère ne peuvent plus être prévus de façon déterministe. L'écoulement atmosphérique à ces échelles est instable et son caractère chaotique fait qu'il ne peut être connu que par ses propriétés statistiques : il est turbulent. Les phénomènes non-linéaires jouent un rôle prépondérant dans un écoulement turbulent, l'énergie étant transférée entre les structures de tailles très diverses présentes dans l'écoulement.

Bien qu'il soit difficile de caractériser la turbulence, ses nombreuses propriétés sont relativement bien connues : la turbulence est instationnaire, dissipative, tri-dimensionnelle et rotationnelle, diffusive. La turbulence favorise donc le mélange : elle augmente les transferts de quantité de mouvement, de chaleur ou de masse de plusieurs ordres de grandeur. Elle permet donc les transferts de chaleur du sol vers l'atmosphère et le mélange des polluants.

#### 1.1.3.5 Le dépôt

Le dépôt est un puits par lequel les polluants quittent l'atmosphère pour la surface du sol, des bâtiments ou de la végétation.

Le dépôt humide a lieu en présence de précipitations qui «lessivent» l'atmosphère en entraînant les polluants vers les surfaces. La grande variabilité des précipitations rend difficile la quantification du dépôt humide ; celui-ci dépend de plus de la solubilité de l'espèce dans la neige ou la pluie (qui varie en fonction de la température et du pH), de la taille des gouttes et de leur nombre (Finlayson-Pitts et Pitts 2000).

Le dépôt sec (en l'absence de précipitations) dépend principalement du degré de turbulence de l'atmosphère, des caractéristiques de la surface et des propriétés chimiques des polluants en contact avec celle-ci. Son mécanisme peut être décomposé en trois étapes : (i) une étape aérodynamique qui est le transfert du gaz vers la couche laminaire c'est-à-dire celle qui est en contact avec la surface, (ii) une étape de diffusion, à travers la couche laminaire et (iii) une étape de piégeage du gaz par la surface. La vitesse de dépôt d'un gaz sur une surface est souvent estimée grâce à un modèle analogue à celui de résistance électrique, les résistances correspondant aux trois étapes du processus étant en série (Wesely 1989; Finlayson-Pitts et Pitts 2000).

## **1.1.4 La mesure de la pollution**

### **1.1.4.1 Moyens et objectifs**

#### **Les réseaux**

Depuis l'an 2000, le dispositif de surveillance de la qualité de l'air et d'information, défini dans la LAURE, est géré sur le terrain par une quarantaine d'Associations Agréées pour la Surveillance de la Qualité de l'Air (AASQA), regroupées au sein de la Fédération ATMO. L'objectif de ce dispositif de réseaux est triple : surveiller la qualité de l'air pour informer les autorités et les citoyens mais aussi comprendre et prévoir les épisodes de pollution et évaluer les impacts des mesures de prévention et réduction de la pollution.

Les treize polluants réglementés, qui doivent être pris en compte dans le cadre de l'évaluation et de la gestion de la qualité de l'air dans l'Union Européenne, comptent notamment l'anhydride sulfureux, le dioxyde d'azote, le monoxyde de carbone, les particules fines (notamment les PM<sub>10</sub>, pour Particulate Matter d'un diamètre inférieur à 10 microns) et les particules en suspension et l'ozone ainsi que le benzène, les hydrocarbures aromatiques polycycliques (HAP) et des métaux (le plomb, le cadmium, l'arsenic, le nickel et le mercure) (Directive 96/62/CE du Conseil du 27 septembre 1996 concernant l'évaluation et la gestion de la qualité de l'air ambiant, Journal officiel n° L 296 du 21/11/1996 p. 0055 - 0063).

#### **Les campagnes de mesure**

Dans le cadre de la recherche, des campagnes de mesure sont organisées. Elles consistent en la réalisation de nombreuses mesures pendant une période de temps réduite. La mobilisation d'une grande quantité de matériel (et d'un personnel nombreux) permet d'effectuer des mesures originales par rapport aux réseaux, par exemple en-dehors des zones couvertes ou pour des espèces non mesurées en routine. L'Annexe B indique les principales caractéristiques de la campagne ESQUIF qui constitue un bon exemple de campagne complète et sera utilisée pour ce travail. De telles campagnes sont, bien sûr, particulièrement coûteuses. Leur but est de récolter un maximum d'informations qui serviront notamment à valider les modèles.

### 1.1.4.2 Intérêt et difficultés

Les mesures constituent une information directe sur le système réel, c'est pourquoi elles permettent notamment la validation des modèles (voir page 24). Elles ont cependant deux limitations : l'information est discontinue (en espace et en temps) et nécessairement incomplète (en nombre d'espèces par exemple, pour des raisons techniques ou financières). Toute mesure comporte, en plus de l'incertitude liée à l'instrument, une erreur dite **de représentativité**. Une mesure est en effet représentative de la réalité dans un certain volume autour du point de mesure. Or, ce volume dépend de phénomènes locaux (dynamique, émissions par exemple) et de l'emplacement de la station. En milieu urbain par exemple, une mesure d'ozone peut n'être représentative que des quelques mètres de trottoir qui l'entourent puisque les émissions de NO des voitures titrent l'ozone au-dessus de la rue. En milieu rural ou en altitude, la représentativité d'une mesure d'ozone peut s'étendre à quelques kilomètres. Pour un polluant à courte durée de vie comme NO, qui disparaît donc très vite lorsque la masse d'air s'éloigne de sa source, la question de la représentativité des mesures est encore plus aigüe.

Le problème de la représentativité des mesures de surface est en partie pris en compte par la classification européenne des stations en six types, selon leur environnement :

- les stations rurales, situées loin des agglomérations permettent de suivre les transferts de pollution et les polluants secondaires produits loin des sources comme l'ozone,
- les stations urbaines et péri-urbaines, situées loin des sources de polluants primaires ont pour but d'estimer la pollution à laquelle est soumise la population au cours de toute sa journée,
- les stations «trafic», situées au plus près de la circulation pour évaluer l'exposition maximale potentielle des piétons et automobilistes,
- les stations industrielles, situées près d'installations particulières (usines, raffineries par exemple),
- les stations d'observation, qui ne répondent pas aux critères précédents et sont utilisées pour compléter les autres informations dans un cadre de recherche.

Le réseau AIRPARIF est décrit page 86.

## 1.1.5 Études, prévisions, scénarios et modèles numériques

### 1.1.5.1 Nécessité de la modélisation numérique

Quatre types de besoins ont amené au développement des modèles numériques **déterministe de chimie-transport** :

- le besoin de comprendre des phénomènes physiques complexes et particulièrement de quantifier le rôle de chacun (par exemple la part due à l'importation d'ozone et celle de la production locale). La non-linéarité du système chimique notamment rend les interactions entre les différents phénomènes particulièrement difficiles à interpréter.
- le besoin de reconstituer des informations inaccessibles : estimer les concentrations aux endroits ou aux moments où il n'y a pas de mesure ; reconstituer l'évolution des concentrations dans le passé.
- le besoin de prévoir : la qualité de l'air pour prendre des mesures, avertir la population.
- le besoin d'évaluer des scénarios : évaluer l'impact de mesures de réduction des émissions ou l'impact de la non-application de ces mesures aujourd'hui et dans l'avenir.

### 1.1.5.2 Ce qu'est un modèle déterministe

Un modèle est un ensemble d'équations décrivant une certaine réalité physique. À partir d'un certain nombre de données ou **paramètres** ou entrées, les équations, choisies pour décrire les lois physiques, ou les paramétrisations, choisies pour décrire des relations empiriques, sont intégrées pour obtenir les **prévisions** c'est-à-dire les valeurs des sorties. Dans le cas de la pollution atmosphérique à l'échelle régionale, les paramètres comprennent entre autres le champ de vent, le profil vertical de température (les variables météorologiques sont des paramètres : les modèles sont dits **off-line**, par opposition aux modèles «on-line» qui calculent également ces variables), la topographie du domaine, le cadastre d'émission, ... (Figure 1.5). Les prévisions sont les champs de concentrations des polluants d'intérêt pour l'étude, par exemple l'ozone pour les pics estivaux.

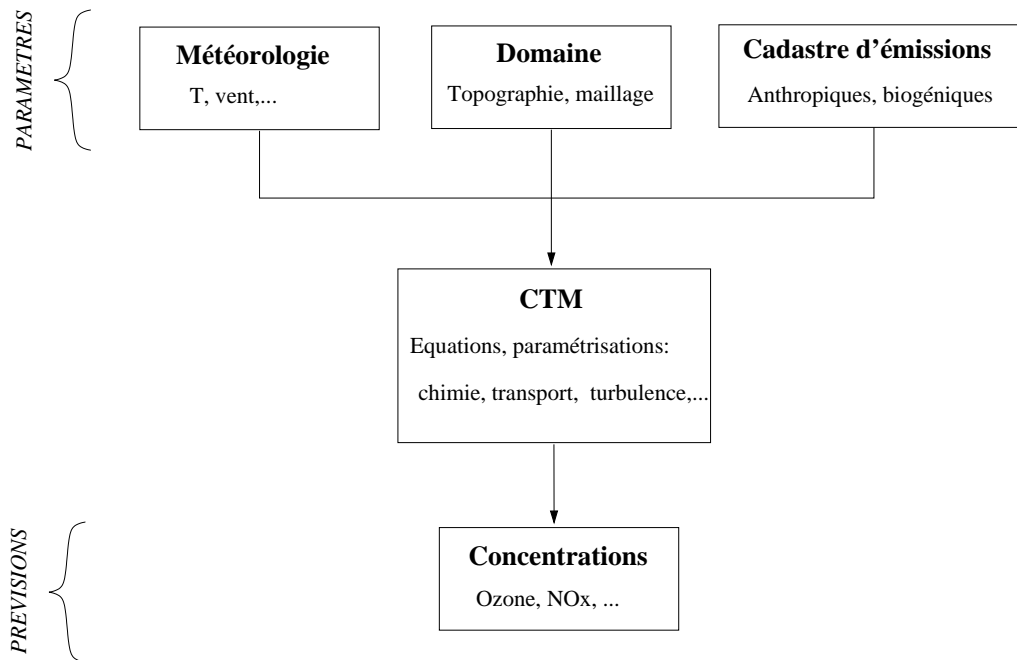


FIG. 1.5 – Schéma de principe d'un modèle de chimie-transport (CTM) déterministe.

Un modèle est un système discret car on choisit un nombre limité d'équations et de variables et on les décompose en espace et en temps. Les choix effectués parmi toutes les équations et les variables possibles doivent être adaptés au but poursuivi : il n'y a pas «le bon modèle» mais «un bon modèle pour...» À l'échelle régionale, les modèles de chimie-transport le plus souvent utilisés sont de type **eulérien** (EUROTRAC-2 2001) : un maillage est placé sur le domaine et la quantité de polluant qui transite dans chaque boîte ainsi délimitée est calculée.

### 1.1.5.3 Avantages et limites des modèles

Les modèles numériques comportent des avantages qui en font des outils complémentaires des mesures : ils intègrent l'ensemble des connaissances pour fournir une information continue et potentiellement très complète. Ils permettent de quantifier les interactions entre les différents processus.



Ils se heurtent cependant à trois limites. La première est due au fait qu'un modèle ne peut justement délivrer *que* l'état des connaissances ; il ne peut évidemment prendre en compte, par exemple, une réaction chimique inconnue. Le deuxième obstacle est dû à la quantité d'informations nécessaire en entrée d'un modèle de chimie-transport, informations de type météorologique, géophysique, chimique,... Enfin, un modèle n'est qu'une représentation de la réalité et doit donc être validé, ceci pour garantir que la représentation est assez fidèle pour l'usage auquel elle est destinée (prévision ou recherche par exemple). Cette validation est peut-être l'étape la plus délicate de l'élaboration d'un modèle ; elle est même, en toute rigueur, impossible.

La validation est la procédure permettant de jauger la pertinence des résultats simulés. Il est donc possible de valider une simulation particulière (par exemple, une journée avec un pic d'ozone). Pour cela, différents outils statistiques, de type biais ou erreurs, ont été définis pour comparer les résultats simulés aux mesures. Cette comparaison doit bien sûr prendre en compte de nombreux paramètres dont les plus importants sont les erreurs sur les mesures elles-mêmes (liées à l'instrument) et la différence de représentativité (particulièrement spatiale, parfois aussi temporelle) entre les mesures et le modèle.

## 1.2 Les émissions surfaciques

Comme indiqué page 19, les émissions sont la source des polluants primaires et elles font partie des paramètres d'entrée indispensables aux modèles (page 24).

### 1.2.1 Émissions anthropiques et biogéniques

Les émissions peuvent être réparties en deux familles principales, celle des émissions **anthropiques** et celle des émissions **biogéniques**.

#### 1.2.1.1 Les émissions anthropiques

Les émissions anthropiques sont celles qui sont dues à des activités uniquement humaines. Elles sont particulièrement importantes en milieu urbain où la plus grande part a pour origine les transports (notamment le trafic automobile) et les industries mais aussi le chauffage en hiver. Ce sont les sources d'émissions majoritaires en intensité même si leur emprise spatiale est relativement limitée. Les principaux polluants gazeux émis par l'utilisation de combustibles fossiles sont le dioxyde de soufre ( $\text{SO}_2$ ), les oxydes d'azote ( $\text{NO}$ ,  $\text{NO}_2$ ,  $\text{N}_2\text{O}$ ), le monoxyde de carbone ( $\text{CO}$ ) et des COV. De nombreux COV sont également émis par l'utilisation de solvants et les industries qui les produisent.

Actuellement, d'importants efforts sont réalisés pour mieux quantifier ces émissions, par définition hautement variables en temps (le trafic, par exemple, n'est pas un processus régulier et Kühlwein et Friedrich (2000) indiquent ainsi une erreur potentielle de l'ordre de 35%) et en espace (de nombreuses sources ont un impact très local).

#### 1.2.1.2 Les émissions biogéniques

Les émissions biogéniques (ou naturelles) sont produites par des phénomènes biologiques naturels, éventuellement dans le cadre d'une utilisation humaine comme c'est le cas pour

les émissions des plantes cultivées ou des animaux d'élevage. Les végétaux émettent des COV très réactifs, principalement de l'isoprène et des terpènes. Les bactéries présentes dans le sol émettent entre autres des oxydes d'azote, de l'ammoniac, des composés soufrés et du méthane (CH<sub>4</sub>) (Finlayson-Pitts et Pitts 2000).

En dehors de milieux fortement urbanisés, les émissions biogéniques deviennent la principale source en intensité. Or, ces émissions sont aussi mal connues : l'incertitude sur les totaux annuels d'une base de données d'émissions d'isoprène et de terpènes à l'échelle de l'Europe est estimée à un facteur 2 à 3 (Simpson 1999). Cette grande incertitude est due à la très importante variabilité des émissions d'une espèce à l'autre voire d'une plante à l'autre. De plus, la variation de ces émissions en fonction de facteurs climatiques (température moyenne, pluviométrie,...) reste encore mal comprise. Les processus micro-biologiques fondamentaux de l'émission de composés azotés par les sols et leur dépendance à l'état de ceux-ci ne sont pas encore bien compris non plus. De plus, les espèces chimiques émises ont souvent des temps de vie assez courts ce qui rend les mesures spatialement peu représentatives ; certaines espèces, comme les carbonylés sont difficiles à mesurer. Enfin, isoler dans la concentration mesurée la part issue de l'émission reste délicat.

### 1.2.2 La construction de cadastres «bottom-up»

À l'échelle régionale, les cadastres d'émissions sont généralement construits en suivant une approche dite «bottom-up» : à partir de mesures directes et de statistiques concernant les différents secteurs d'activité, des masses annuelles de NO<sub>x</sub> et de COV sont réparties sur une grille régulière avec une résolution temporelle typique d'une heure (voir [www.emep.int](http://www.emep.int) et [webdab.emep.int](http://webdab.emep.int) pour une description des travaux effectués en Europe). Cette approche est la plus réaliste mais pose un certain nombre de problèmes qui sont actuellement que :

- les mesures ne peuvent décrire toutes les espèces émises (on ne sait pas mesurer certaines espèces, d'autres sont émises en quantité très faible et donc difficilement détectable).
- les statistiques ne sont pas toujours précises. Les émissions anthropiques issues de l'industrie, par exemple, sont estimées statistiquement à partir de déclarations d'activités industrielles qui ne sont pas de manière sûre exactes et à jour.
- les émissions changent en permanence et il est difficile d'actualiser les masses émises (Kühlwein et Friedrich 2000) ; ces changements se produisent à différentes échelles de temps, de l'heure à l'année, tant pour les émissions biogéniques (en fonction de la lumière, des précipitations, de la couverture végétale,...) que pour les émissions anthropiques (en fonction du trafic, de la production industrielle, de facteurs socio-économiques). Le problème se pose de façon aiguë en milieu urbain, où les émissions sont en grande partie issues d'activités comme la circulation automobile ou les petites industries qu'il est difficile d'estimer et, surtout, de suivre.

Finalement, la composition des cadastres d'émissions est complexe : les données «brutes» permettant de les constituer sont souvent incomplètes et insuffisantes. Comme elles sont hétérogènes en temps et en espace, il est difficile de regrouper ces nombreuses informations en un ensemble cohérent.

### 1.2.3 La mise en œuvre dans les modèles : les espèces-modèle

Il est impossible, dans un modèle, de prendre en compte toutes les espèces réellement présentes dans l'atmosphère et toutes les réactions possibles. Cela est dû non seulement à leur nombre élevé mais aussi à un manque de connaissance certain. Les modèles utilisent des espèces-modèle. Elles correspondent à des groupements d'espèces réelles ayant des propriétés réactives (supposées) proches. La chimie de l'ensemble des espèces réelles est alors représentée par celle de l'espèce-modèle choisie. Les cadastres d'émissions doivent donc être exprimés en fonction de ces espèces-modèle, du moins celles qui correspondent à des polluants primaires.

Dans un cadastre d'émissions, certaines espèces (réelles) sont corrélées. C'est le cas par exemple de NO et NO<sub>2</sub>, dont les émissions ne pouvaient être mesurées indépendamment. Les cadastres ont donc été élaborés à partir de mesures de concentrations de NO<sub>x</sub> effectuées assez près des sources pour être considérées comme dues uniquement aux émissions. Cependant, la chimie modélisée a besoin de connaître les émissions de NO et de NO<sub>2</sub> séparément puisque chacune des espèces correspond à une espèce-modèle différente, subissant des réactions chimiques différentes. Les émissions de NO<sub>x</sub> du cadastre «brut» sont alors arbitrairement réparties en NO et NO<sub>2</sub> grâce à un coefficient empirique (par exemple 90% de NO et 10% de NO<sub>2</sub> pour des émissions dues au trafic routier). Selon le degré de détail du cadastre «brut» et les espèces-modèle utilisées, ce type d'hypothèses concerne un plus ou moins grand nombre d'espèces.

## 1.3 Les objectifs de cette thèse

Pour les modèles de chimie-transport déterministes utilisés en pollution atmosphérique (page 23), les émissions sont un paramètre d'entrée indispensable auquel les concentrations simulées sont très sensibles. Dans le cas de la pollution photochimique/photo-oxydante à l'échelle régionale, le rôle des émissions est crucial : la qualité des concentrations simulées est directement liée à celle du cadastre d'émissions utilisé (Menut 2003). Une meilleure détermination des flux d'émissions permettrait donc de mieux simuler les champs de concentrations et de mieux comprendre les phénomènes de pollution. Cependant, il semble irréaliste de revoir les méthodes de construction des cadastres (page 26) pour réduire l'incertitude. Il sera donc **toujours impossible de disposer de données d'émissions à jour et à la résolution spatiale et temporelle voulue**.

Il reste donc la possibilité de **modifier les cadastres existant** en y ajoutant de l'information permettant d'obtenir des cadastres à jour. Cette information existe : elle est disponible dans les mesures de concentrations d'espèces chimiques effectuées en surface (et parfois en altitude) par les réseaux ou lors de campagnes de mesures.

**Le but de l'optimisation des émissions** est donc de trouver le **cadastre optimal** c'est-à-dire **celui qui minimise la différence entre les champs de concentrations simulés et mesurés**. Le résultat obtenu se présente alors sous la forme de **facteurs correctifs pour les flux de chaque espèce** (ou famille) émise.

Il s'agit donc de retrouver les émissions «qu'il aurait fallu donner en entrée au modèle» pour obtenir des concentrations simulées en sortie du modèle égales aux concentrations mesurées (ou presque égales : on traitera plus tard du problème de la qualité des mesures et de leur représentativité par rapport au modèle, page 38).

Un tel problème est dit **inverse** : il s'agit de trouver les valeurs de paramètres du modèle **direct**, en l'occurrence des flux d'émission, à partir d'**observations**, ici les mesures de concentrations, qui apportent de l'information sur les prévisions, c'est-à-dire les concentrations simulées.

Les méthodes permettant de résoudre de tels problèmes sont (logiquement) appelées **méthodes inverses** et sont décrites à la Section 2.1.1 page 29.

Le premier objectif de cette thèse est de mettre au point une méthodologie d'inversion des émissions anthropiques de  $\text{NO}_x$  et de COV à l'échelle régionale dans les grandes agglomérations et leur région (comme Paris et l'Île-de-France), en utilisant les mesures de routine des réseaux ( $\text{O}_3$  et  $\text{NO}_x$ ). Ce dernier point doit permettre d'une part de mener des études lors d'épisodes passés variés et, d'autre part, d'envisager une utilisation de l'inversion pour l'amélioration des prévisions et finalement d'assurer la validité de la méthode pour une large gamme de périodes et de lieux.

Une fois validée, la méthodologie élaborée sera appliquée en Île-de-France

## 1.4 Plan du manuscrit

Le chapitre 2 propose un état de l'art des méthodes inverses, une analyse des possibilités et des limites de ces méthodes à l'échelle régionale et les choix effectués pour notre travail, notamment l'utilisation d'une méthode de krigeage couplée à l'inverse.

Le chapitre 3 présente le modèle CHIMERE et son adjoint, l'inverse élaboré sur ces bases et enfin, une validation du code développé grâce à des cas académiques simples.

Le chapitre 4 détaille l'approche originale utilisant le krigeage et présente les résultats obtenus sur des cas académiques réalistes représentant l'Île-de-France destinés à valider cette méthodologie.

Le chapitre 5 complète la méthode élaborée par une préparation en zones qui a pour but de réduire la dimension des problèmes réels. La méthode complète est validée sur des cas académiques.

Le chapitre 6 expose les résultats obtenus pour l'optimisation des émissions en Île-de-France en été.

Enfin, la dernière partie résume les conclusions de ce travail et indique des perspectives.

# Chapitre 2

## Méthodes inverses

### 2.1 État de l’art

#### 2.1.1 Principe général et historique

Comme indiqué à la Section 1.3 page 27, le but des méthodes inverses est de résoudre le problème inverse c’est-à-dire d’estimer, à partir d’observations, les valeurs des paramètres du modèle permettant de retrouver les prévisions souhaitées.

En pratique, les méthodes inverses consistent à minimiser la différence entre l’observation et la prévision pour une variable (ici les concentrations de polluants) afin de mieux estimer certains paramètres (les émissions par exemple) dont dépend celle-ci. La méthode d’inversion modifie les valeurs de départ des paramètres, dites *a priori* (ou «first-guess») pour obtenir les valeurs **optimisées**. On appelle **dimension du problème** (inverse) le nombre de paramètres à inverser, que l’on regroupe en un **vecteur d’état**. Dans le cas des émissions par exemple, chaque flux de chaque espèce émise constitue un paramètre d’entrée.

On appelle **contraintes** les observations c’est-à-dire les valeurs que les prévisions doivent approcher. Des mesures de concentration sont classiquement utilisées comme contraintes pour l’inversion des émissions. Pour comparer les valeurs des contraintes aux prévisions, on relie le vecteur d’état à l’espace des observations par un **opérateur d’observation**. Celui-ci n’est autre que le modèle qui permet de calculer les valeurs simulées aux mêmes points que les observations : par exemple, un modèle de chimie-transport qui calcule les concentrations en fonction des émissions puis un système d’interpolation permettant d’obtenir les concentrations simulées aux emplacements des stations de mesure (en général, les points de grille du modèle et les emplacements des mesures ne coïncident pas).

La modélisation inverse fait donc partie de la famille des techniques d’**hybridation entre mesures et modèle** dont le but commun est de définir au mieux l’état d’un système en utilisant le maximum d’informations.

Les premières applications de ces techniques avec des **modèles déterministes eulériens** ont consisté en le développement de l’**assimilation de données** pour l’océanologie et la météorologie (voir Daley (1996a) et Daley (1996b) pour une revue historique). L’assimilation de données consiste à intégrer des informations telles que des observations pour améliorer l’**évolution** du modèle. Il s’agissait alors de traiter de grandes échelles spatiales sur de longues périodes (de type «climatologies»). L’application de ce type de techniques à des problèmes concernant les gaz-trace de l’atmosphère est plus récente et porte la trace

de cette histoire : Elbern et al. (1997) présentent l'assimilation de données chimiques dans la troposphère à l'échelle globale et l'ouvrage de Kasibhatla et al. (2000) permet de faire le lien entre l'atmosphère et l'océan, toujours à l'échelle globale. Parmi les nombreuses classes d'algorithmes qui ont été développées peu à peu, les plus couramment employées aujourd'hui sont l'**assimilation séquentielle** (qui prend en compte les observations les unes après les autres, dans l'ordre chronologique) et l'**assimilation variationnelle** (qui prend en compte les observations dans une **fenêtre temporelle** d'une durée adaptée au problème à traiter). Leur théorie est décrite de manière exhaustive dans Talagrand (1997). Bien que le but de l'inversion, qui est d'estimer des **paramètres** du modèle, soit différent de celui de l'assimilation de données, les algorithmes développés pour cette dernière ont pu servir de base aux techniques d'inversion proprement dite (Ménard et al. 2004). L'ouvrage de Enting (2002) est consacré aux problèmes inverses rencontrés dans l'atmosphère et Reeves et al. (2004) traitent de l'inversion des émissions à partir de mesures de concentrations. C'est ce dernier point qui nous intéresse dans le cadre de ce travail et on présente donc, dans ce qui suit, un état de l'art de l'inversion des émissions de gaz-trace dans l'atmosphère à l'aide de modèles eulériens.

### 2.1.2 Quelles sont les méthodes utilisées ?

Les travaux réalisés jusqu'à présent portent principalement sur l'échelle globale ou continentale et utilisent plus souvent la méthode du **filtre de Kalman** (Hartley et Prinn 1993; Haas-Laursen et al. 1996; Hein et al. 1997; Mahowald et al. 1997; Jacob et al. 2002; Wang et Bentley 2002; Gilliland et al. 2003) que celle de l'**adjoint** (Houweling et al. 1999; Müller et Stavrakou 2005). À l'échelle régionale, trois des huit études disponibles utilisent le filtre de Kalman (Mulholland et Seinfeld 1995; Chang et al. 1996; Chang et al. 1997) et quatre, la méthode de l'adjoint (Elbern et al. 1997; Elbern et Schmidt 1999; Elbern et al. 2000; Quélo et al. 2005).

#### Le filtre de Kalman

Originellement, cette méthode est **séquentielle** et basée sur la relation linéaire entre la modification des contraintes (des observations) et les paramètres pendant la même période (Hartley et Prinn 1993). Le but du filtre de Kalman est d'obtenir, à un instant donné, la meilleure estimation possible du vecteur d'état compte-tenu des observations disponibles à cet instant. Le principe de l'algorithme, basé sur la récurrence, est le suivant (Kalman 1960; Kalman et Bucy 1961; Lacarra et Talagrand 1988; Enting 2002) :

- 1- On dispose d'un modèle permettant :
  - de relier, à n'importe quel instant  $t_n$  le vecteur d'état  $x(t_n)$  aux variables observées  $y(t_n)$  grâce à l'opérateur d'observation  $\mathbf{H}_n$  selon :  $y_{sim}(t_n) = \mathbf{H}_n x(t_n)$ . .
  - de prévoir le vecteur d'état  $x^f(t_{n+1})$  à l'instant  $t_{n+1}$  ( $x^f(t_{n+1})$  est appelé prévision, <sup>f</sup> pour «forecast») à partir de l'état  $x^a(t_n)$ , analyse obtenue à l'instant précédent, soit :  $x^f(t_{n+1}) = \mathbf{M}_{n \rightarrow n+1} x^a(t_n)$ , le modèle  $\mathbf{M}_{n \rightarrow n+1}$  étant valable entre les instants  $t_n$  et  $t_{n+1}$ .
- 2- Les covariances d'erreur de la prévision  $x^f(t_{n+1})$ , contenues dans  $\mathbf{P}_{n+1}^f$ , dépendent alors des covariances d'erreur de l'analyse  $\mathbf{P}_n^a$  et des covariances d'erreur du modèle correspondant à  $\mathbf{M}_{n \rightarrow n+1}$ , contenues dans  $\mathbf{Q}_n$ . On a :  $\mathbf{P}_{n+1}^f = \mathbf{M}_{n \rightarrow n+1} \mathbf{P}_n^a \mathbf{M}_{n \rightarrow n+1}^T +$

$\mathbf{Q}_n$ . On remarquera que l'on suppose que les erreurs du modèle et les erreurs de l'analyse ne sont pas corrélées.

- 3- Le gain **optimal**  $\mathbf{K}_{n+1}$ , qui contient les poids permettant d'effectuer l'analyse au temps  $t_{n+1}$ , dépend des covariances d'erreur de la prévision  $\mathbf{P}_{n+1}^f$  et des covariances d'erreur d'observation correspondant à  $\mathbf{H}_{n+1}$ , contenues dans  $\mathbf{R}_{n+1}$ . Il est donné par :  $\mathbf{K}_{n+1} = \mathbf{P}_{n+1}^f \mathbf{H}_{n+1}^T [\mathbf{H}_{n+1} \mathbf{P}_{n+1}^f \mathbf{H}_{n+1}^T + \mathbf{R}_{n+1}]^{-1}$ .  $\mathbf{P}_{n+1}^f \mathbf{H}_{n+1}^T$  contient donc les covariances d'erreur du modèle entre les points de grille et les points de mesure ;  $\mathbf{H}_{n+1} \mathbf{P}_{n+1}^f \mathbf{H}_{n+1}^T$  réduit  $\mathbf{P}_{n+1}^f$  aux seuls points de mesure.
- 4- Une fois les poids calculés dans  $\mathbf{K}_{n+1}$ , on peut effectuer l'analyse de l'état au temps  $t_{n+1}$ , qui correspond à la prévision effectuée à la première étape « corrigée » par les valeurs utilisables au temps  $t_{n+1}$  contenues dans le vecteur  $y_{obs}(t_{n+1})$ . On a :  $x^a(t_{n+1}) = x^f(t_{n+1}) + \mathbf{K}_{n+1} [y_{obs}(t_{n+1}) - \mathbf{H}_{n+1} x^f(t_{n+1})]$ . Le vecteur  $y_{obs}(t_{n+1}) - \mathbf{H}_{n+1} x^f(t_{n+1})$  contient simplement les écarts entre valeurs mesurées et valeurs simulées  $y_{obs}(t_{n+1}) - y_{sim}(t_{n+1})$ .
- 5- Enfin, on obtient les covariances d'erreur de la nouvelle analyse  $\mathbf{P}_{n+1}^a$  grâce à :  $\mathbf{P}_{n+1}^a = [\mathbf{I} - \mathbf{K}_{n+1} \mathbf{H}_{n+1}] \mathbf{P}_{n+1}^f$ .

Il est nécessaire d'initialiser la procédure avec  $x^a(t_0)$  et  $\mathbf{P}_0^a$

Les problèmes rencontrés en océanologie et météorologie étant en général non-linéaires, le filtre de Kalman dit étendu généralise l'algorithme à des modèles et des opérateurs d'observation non-linéaires en les linéarisant au voisinage de l'analyse à effectuer. Pour les problèmes fortement non-linéaires, il est possible d'utiliser le filtre de Kalman dit d'ensemble, introduit par Evensen (1994) et utilisant les méthodes stochastiques. Enfin, le lissage de Kalman (« Kalman smoother » en anglais) permet d'étendre le filtre de Kalman à une fenêtre temporelle (Gelb 1974).

### L'approche adjointe

Le principe de cette méthode **variationnelle** est résumé sur la Figure 2.1. À partir des forçages de la météo et des émissions, le modèle de chimie-transport direct (à gauche) fournit des champs de concentrations que l'on peut comparer aux observations effectuées. La modélisation inverse (représentée à droite) consiste à minimiser la différence entre les concentrations simulées et les concentrations observées en modifiant les émissions. La **distance** entre les concentrations simulées et les contraintes est représentée par une **fonction coût**, qui doit donc être minimisée. La fonction coût bien choisie est calculée par le modèle direct alors que son gradient par rapport aux paramètres (par exemple, les émissions) est fourni par l'adjoint. Celui-ci permet de calculer la dérivée d'une quantité simulée (par exemple, la concentration d'une espèce donnée dans une maille donnée) à tous les paramètres d'entrée (par exemple, tous les flux d'émissions). À partir de la valeur du coût et de son gradient en un point, un minimiseur permet de trouver la correction à apporter aux paramètres pour se rapprocher des observations et répète le cycle direct-adjoint jusqu'à obtenir d'une part un écart minimal entre variables simulées et observées et d'autre part les corrections qui lui correspondent c'est-à-dire l'ensemble de paramètres (par exemple le cadastre d'émissions) qui donne les meilleurs résultats de simulation.

Il existe plusieurs choix de formulation du coût, adaptés aux différents problèmes traités. Cette fonction du vecteur d'état permet en général d'indiquer d'une part, à quel point on souhaite se rapprocher des contraintes et d'autre part, à quel point on peut s'éloigner des

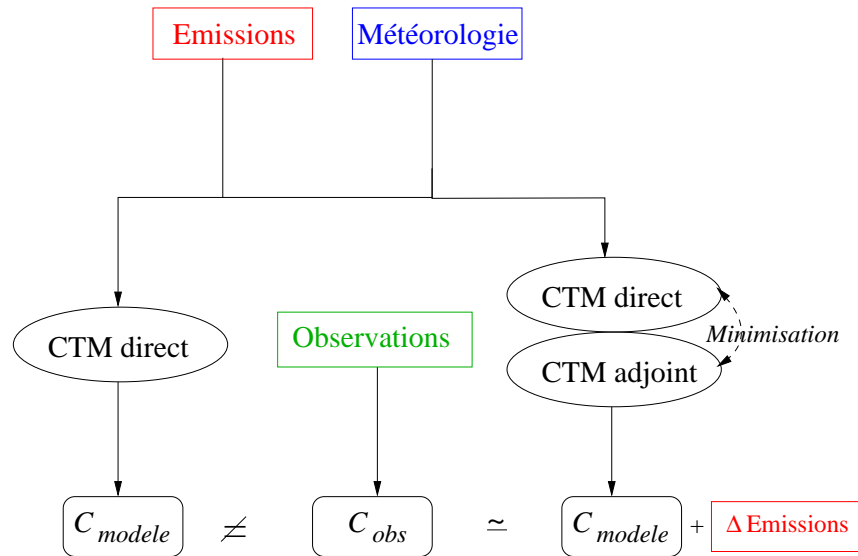


FIG. 2.1 – Principe de la modélisation inverse. CTM = Chemistry Transport Model

paramètres *a priori*. Le coût est donc exprimé le plus souvent à l'aide de deux vecteurs contenant, pour l'un, les différences entre les concentrations simulées et les contraintes correspondant, pour l'autre, les différences entre les paramètres optimisés et *a priori*. Elle prend alors la forme (Talagrand 1997; Elbern et al. 1997; Elbern et Schmidt 1999; Houweling et al. 1999; Elbern et al. 2000; Daley 1997; Prinn et Hartley 1995) :

$$J(x(t)) = (y_{obs}(t) - \mathbf{H}x(t))^T \underbrace{\mathbf{R}}_{y(t)} (y_{obs}(t) - \mathbf{H}x(t)) + (x_{fg}(t) - x(t))^T \underbrace{\mathbf{B}}_{X(t)} (x_{fg}(t) - x(t))$$

où :

- $x(t)$  est le vecteur d'état ;  $t$  n'indique pas un instant dans le temps mais un état de l'optimisation
- $\mathbf{H}$  est l'opérateur d'observation
- $y(t)$  est le vecteur qui contient les différences entre les concentrations simulées ( $y_{obs}(t)$ ) et les contraintes correspondant ( $y_{sim}(t) = \mathbf{H}x(t)$ )
- $X(t)$  est le vecteur qui contient les différences entre les paramètres optimisés ( $x(t)$ ) et *a priori* ( $x_{fg}(t)$ ,  $fg$  pour «first-guess» ou ébauche)
- $\mathbf{R}$  est la matrice des **covariances d'erreur d'observation** Ses termes diagonaux correspondent directement aux variances et les termes extra-diagonaux représentent les éventuelles corrélations entre ces erreurs, qui peuvent être importantes par exemple dans le cas de mesures effectuées avec le même instrument, comme à bord d'un satellite.
- $\mathbf{B}$  est la matrice des **covariances d'erreur du modèle** et est souvent appelée matrice «**de background**». Ses termes diagonaux sont des poids qui s'appliquent aux composantes du vecteur des différences entre paramètres optimisés et *a priori*. Ils permettent donc de pénaliser les solutions de l'optimisation qui s'éloignent trop de l'*a priori* en prenant en compte son incertitude. Ceci est particulièrement utile dans les cas où cette incertitude n'est pas trop élevée. Les termes extra-diagonaux de la matrice «de back-



ground» permettent d'ajouter d'autres informations *a priori* dans l'espace des paramètres à inverser telles que :

- des liens temporels. Ceci est particulièrement utile en météorologie puisque l'on traite des phénomènes continus. Si une erreur sur la température est constatée à l'heure  $t$ , la même erreur est très probablement présente à  $t + \Delta t$ .
- des liens spatiaux. Les termes de la matrice établissent une corrélation spatiale entre les espèces émises au même moment dans tout le domaine. À l'échelle continentale, des fonctions de corrélation peuvent être élaborées pour représenter une zone cohérente comme un désert par exemple.

### 2.1.3 Quelles émissions ont été inversées ?

Les plus récentes des études d'inversion d'émissions effectuées avec des modèles de chimie-transport eulériens tridimensionnels sont indiquées dans le Tableau 2.1. Les méthodes inverses ont principalement été utilisées à l'échelle **globale** (couvrant toute la planète) ou **continentale**. Les espèces d'intérêt à ces échelles spatiales sont celles qui ont des temps de vie longs (voir Figure 1.2) comme le méthane (Hein et al. 1997; Houweling et al. 1999; Wang et Bentley 2002), le dioxyde de carbone (Bousquet et al. 1999a; Kaminski et Heimann 2001; Rödenbeck et al. 2003; Gurney et al. 2005), les CFC (Hartley et Prinn 1993; Mahowald et al. 1997) ou encore le monoxyde de carbone (Bergamaschi et al. 2000; Pétron et al. 2002; Pétron et al. 2004; Müller et Stavrakou 2005; Wang et al. 2004). En plus de l'inversion des sources d'émissions, certaines études se préoccupent des puits comme Bousquet et al. (1999a) pour le dioxyde de carbone, voire du bilan général de l'espèce comme Jacob et al. (2002) pour l'acétone. Des cas-tests permettent d'aborder le problème de la qualité et de l'éventuelle amélioration du réseau de mesure pour l'inversion (Hartley et Prinn 1993; Mahowald et al. 1997).

La question de la **variabilité temporelle** des sources est particulièrement détaillée par Haas-Laursen et al. (1996), Hein et al. (1997), Pétron et al. (2004) et Gurney et al. (2005). Bousquet et al. (1999a) traitent une année représentative de la climatologie de toute une décennie. Les aspects théoriques de l'inversion de flux d'émission variables dans le temps sont traités pour le dioxyde de carbone à l'échelle globale par Baker (2000).

Les corrections apportées aux cadastres *a priori* peuvent aussi présenter une **variabilité spatiale** propre. À l'échelle globale, Müller et Stavrakou (2005) obtient une augmentation des émissions de CO de 28% et une diminution des émissions de NO<sub>x</sub> de 14% mais cette dernière correspond à une augmentation des flux tropicaux et à une diminution en Europe et en Asie. À l'échelle continentale, Wang et al. (2004) obtiennent, pour la Chine, des augmentations moyennes des flux de CO et des NO<sub>x</sub> de 43 et 47% respectivement; la répartition spatiale des corrections à travers le pays est très hétérogène mais peut être expliqué de façon réaliste.

Enfi, les résultats obtenus consistent non seulement en de nouvelles valeurs pour les sources, mais aussi en une réduction de l'**incertitude** sur celles-ci, qui peut atteindre 75% pour un hémisphère entier (Houweling et al. 1999) ou de 25 à 80% pour un continent (Wang et Bentley 2002).

Les références citées ici ne forment pas une liste exhaustive des études effectuées à l'échelle globale. De nombreuses équipes travaillent actuellement sur l'inversion des émissions des gaz à effet de serre comme le dioxyde de carbone.

Référence	Espèces inversées	Espèces mesurées
Échelle globale		
Hartley et Prinn (1993)	CFC1 <sub>3</sub>	CFC1 <sub>3</sub>
Hein et al. (1997)	CH <sub>4</sub>	CH <sub>4</sub>
Mahowald et al. (1997)	CFC1 <sub>3</sub>	CFC1 <sub>3</sub>
Bousquet et al. (1999a)	CO <sub>2</sub>	CO <sub>2</sub>
Houweling et al. (1999)	CH <sub>4</sub>	CH <sub>4</sub>
Bergamaschi et al. (2000)	CO	CO
Kaminski et Heimann (2001)	CO <sub>2</sub>	CO <sub>2</sub>
Jacob et al. (2002)	acétone	acétone
Pétron et al. (2002)	CO	CO
Rödenbeck et al. (2003)	CO <sub>2</sub>	CO <sub>2</sub>
Pétron et al. (2004)	CO	CO (sat)
Gurney et al. (2005)	CO <sub>2</sub>	CO <sub>2</sub>
Müller et Stavrakou (2005)	CO, NO <sub>x</sub>	CO, NO <sub>2</sub> (sat)
Échelle continentale		
Wang et Bentley (2002)	CH <sub>4</sub>	CH <sub>4</sub>
Gilliland et al. (2003)	NH <sub>3</sub>	NH <sub>4</sub> <sup>+</sup>
Wang et al. (2004)	CO, NO <sub>x</sub>	CO, NO <sub>y</sub>
Échelle régionale		
Mulholland et Seinfeld (1995)	CO	CO
Chang et al. (1996)	isoprène	isoprène
Chang et al. (1997)	CO	CO
Elbern et al. (1997)	étude théorique	
Elbern et Schmidt (1999)	NO <sub>2</sub>	O <sub>3</sub>
Elbern et al. (2000)	NO, COV	O <sub>3</sub>
Mendoza-Dominguez et Russell (2001)	COV, CO, NO <sub>x</sub>	COV, CO, NO, NO <sub>2</sub> , NO <sub>x</sub> , NO <sub>y</sub> , O <sub>3</sub>
Quélo et al. (2005)	NO <sub>x</sub>	NO, NO <sub>2</sub> , O <sub>3</sub>

TAB. 2.1 – Études en modélisation inverse des émissions à l'échelle régionale.

Les études de modélisation inverse à l'**échelle régionale** sont rares (huit sont indiquées dans le Tableau 2.1). La région d'Atlanta, où a eu lieu une campagne de mesures en 1992 a donné lieu à elle seule à la moitié d'entre elles. L'inversion des émissions de monoxyde de carbone par Mulholland et Seinfeld (1995) à Los Angeles conduit les auteurs à relever la prépondérance des corrections dues aux sources mobiles, notamment le trafic routier. Chang et al. (1996) et Chang et al. (1997) ont inversé respectivement les émissions d'isoprène et de monoxyde de carbone à Atlanta. Dans le premier cas, les résultats sont jugés satisfaisants bien qu'amenant à multiplier les émissions par un facteur deux à dix. Cependant, pour le monoxyde de carbone, des oscillations temporelles irréalistes apparaissent. Les auteurs concluent sur la difficulté à estimer des sources de CO, plus localisées et moins homogènes spatialement que l'isoprène. Mendoza-Dominguez et Russell (2001) se sont intéressés aux précurseurs de l'ozone que sont les COV et les NO<sub>x</sub>. À partir de la

campagne menée à Atlanta, les émissions ont été ajustées en distinguant les contributions des sources anthropiques mobiles, ponctuelles, ou surfaciques et biogéniques. Le principal objectif de Elbern et Schmidt (1999) était l'assimilation de mesures d'ozone. Cependant, comme le cadastre disponible n'était pas à jour, la méthode a été adaptée pour optimiser les émissions des précurseurs de l'ozone. Bien que le problème soit mal posé, les résultats sur les concentrations d'ozone sont améliorés de façon significative avec les émissions optimisées. Les études de Elbern et al. (2000) sur des cas académiques ont montré qu'il est possible d'optimiser les émissions de précurseurs de l'ozone dans des sites urbains complexes à condition de disposer d'un grand nombre de contraintes. Par ailleurs, l'étude conclut à une importante sous-estimation des COV tant en milieu urbain qu'autour de la ville, ce qui rejoint les résultats de Chang et al. (1996) pour l'isoprène. Enfin, l'étude la plus récente et la plus proche de notre problématique a été réalisée par Quélo et al. (2005) dans la région de Lille. Le profil temporel des émissions de  $\text{NO}_x$  au cours d'une semaine a été optimisé à l'aide de mesures de routine d'ozone, de monoxyde d'azote et de dioxyde d'azote. La validation des résultats est effectuée d'une part en appliquant les émissions optimisées à d'autres semaines afin de déterminer si les prévisions sont améliorées et d'autre part en étudiant leur sensibilité au cadastre *a priori*, à la situation météorologique et au mélange vertical.

## 2.2 Limitations des méthodes inverses et échelles d'application

Les techniques d'assimilation de données furent originellement développées pour traiter des problèmes météorologiques, qui sont des problèmes de **conditions initiales**. En effet, les équations régissant l'écoulement atmosphérique étant fortement non-linéaires, la qualité des conditions initiales a une influence déterminante sur celle des résultats de l'intégration complète du modèle.

Les problèmes de chimie-transport sont beaucoup moins non-linéaires. Ils sont du type **conditions aux limites**. Les principaux paramètres dont la qualité détermine celle des sorties du modèle sont en effet les sources surfaciques et, dans le cas de la pollution à l'échelle régionale, les concentrations aux limites du domaine étudié, qui peuvent jouer un rôle prépondérant lors de certains épisodes (voir par exemple (Menut et al. 2000) pour le cas de l'Île-de-France). Les théories développées pour la météorologie ne peuvent être appliquées directement à l'étude d'espèces chimiques à long temps de vie à l'échelle globale et sont encore plus difficiles à adapter au cas de la pollution à l'échelle régionale. Nous nous proposons donc de répertorier les limitations des méthodes d'inversion des émissions en général, qui sont principalement dues aux incertitudes et aux erreurs des mesures, du modèle et de ses données d'entrée, puis les différences entre les échelles d'application.

### 2.2.1 Limites dues aux mesures : toutes les informations nécessaires ne sont pas disponibles

Toutes les informations nécessaires ne sont jamais disponibles dans un cas réel, ceci pour des raisons variées. Les informations fournies par les mesures ne sont pas assez nom-

breuses, ne sont pas adaptées à tous les problèmes, intègrent des effets différents, sont bruitées et enfin, ne sont pas indépendantes les unes des autres.

Surtout si l'on souhaite utiliser les mesures de routine de réseaux, la **quantité de mesures disponibles** est bien inférieure à la dimension du problème qui est alors dit **sous-contraint**. Par exemple en Île-de-France, il y a environ 40 stations Airparif mesurant les concentrations d'ozone. Le cadastre *a priori* est constitué de 625 flux surfaciques par espèce émise. Le rapport entre le nombre de contraintes et la dimension du problème pour l'inversion d'une seule espèce est donc d'environ 5%. Il faut aussi noter que les mesures de surface ne fournissent aucune information sur la verticale. L'utilisation de mesures aéroportées (Jacob et al. 2002; Bousquet et al. 1999a; Wang et al. 2004) ou satellitales (Fisher et Lary 1996; Lyster et al. 1997; Pétron et al. 2004; Müller et Stavrakou 2005) ne permet pas de résoudre ce problème : Müller et Stavrakou (2005) indiquent que l'hémisphère Sud n'est pas assez couvert et Bousquet et al. (1999a) estiment qu'il n'y a tout simplement pas assez de contraintes dans certaines zones. Effectuer des simplifications dans la distribution spatiale des émissions peut permettre de compenser ce manque de contraintes (Enting et Mansbridge 1991) : des régions d'émissions regroupant des zones géographiques présentant des similitudes climatiques et économiques sont définies par Bousquet et al. (1999a) pour l'inversion des flux de dioxyde de carbone. Cependant, les informations *a priori* qui sont utilisées pour simplifier le problème ont une influence non négligeable sur les résultats obtenus (Hein et al. 1997; Bousquet et al. 1999a).

La **pertinence de l'information disponible** constitue un deuxième obstacle. Il n'est pas possible d'inverser des émissions qui n'ont pas d'influence sur les mesures disponibles puisqu'on ne dispose alors d'aucune information. Par exemple, inverser des émissions de HONO avec des mesures de  $\text{NO}_y$  peut ne rien donner parce que HONO peut varier du simple au double sans que son influence dépasse l'ordre du pourcent sur les concentrations mesurées. Même lorsque l'espèce étudiée est à la fois la source à optimiser et la contrainte c'est-à-dire l'espèce mesurée et simulée, cela n'empêche pas les contraintes de ne pas toujours être assez pertinentes : Hein et al. (1997) considèrent que les mesures de méthane disponibles à l'échelle globale sont trop peu sensibles aux émissions ; à l'échelle régionale, une étude inversant les émissions de monoxyde de carbone avec des mesures cette même espèce (Chang et al. 1997) a obtenu des résultats physiquement non réalistes. Le problème se pose également si l'on considère l'évolution temporelle des émissions et des concentrations. En effet, les émissions auxquelles les concentrations sont les plus sensibles peuvent être celles d'heures bien antérieures. C'est le cas par exemple des concentrations d'ozone lors du pic de l'après-midi qui sont très sensibles aux émissions de  $\text{NO}_x$  du pic de trafic le matin et peu sensibles à celles de l'heure précédant le pic. Des mesures d'ozone sont utilisées à l'échelle régionale pour l'inversion des émissions d'oxydes d'azote et de COV (précurseurs de l'ozone) par Elbern et Schmidt (1999), Elbern et al. (2000), Mendoza-Dominguez et Russell (2001), et Quélo et al. (2005) : à condition de disposer de suffisamment de mesures (lors d'une campagne par exemple), les résultats obtenus sont jugés cohérents par rapport aux cadastres d'émissions *a priori* et améliorent les concentrations simulées.

La **qualité de l'information** disponible peut n'être pas assez bonne. L'information fournie par une mesure intègre de nombreux effets différents. La concentration mesurée est un bilan de l'influence de plusieurs sources. Il n'est pas possible *a priori* de séparer les différentes contributions. Lorsque les effets de plusieurs sources se superposent en un même

point de mesure, il n'est pas possible de retrouver les informations concernant chaque source sans d'autres mesures pour lesquelles les effets des sources sont découplés. Mahowald et al. (1997) déplorent que les stations de mesure ne soient pas placées plus près des zones sources. Cependant, même pour une mesure située idéalement dans le panache d'une unique source, il peut être impossible de distinguer entre un mauvais placement de la source et une amplitude mal estimée.

La question de l'**indépendance des mesures** est d'autant plus importante que l'échelle est petite. Pour des études à l'échelle globale (Bousquet et al. 1999a; Bousquet et al. 1999b), on peut considérer qu'une mesure de  $\text{CO}_2$  sur un continent est indépendante d'une mesure effectuée sur un autre continent. Mais plus l'échelle est petite, plus les mesures peuvent être liées. À l'échelle régionale, cette information est contenue dans des covariances. Or, si les variances sont (habituellement) fournies avec les mesures, les covariances ne le sont pas.

Enfin, les **instruments de mesure** eux-mêmes ne sont pas parfaits, les mesures comportent incertitudes et erreurs. L'optimisation peut alors aboutir à des corrections qui transfèrent les erreurs de mesures sur les émissions.

### 2.2.2 Limites dues au modèle : les erreurs dues au modèle n'ont pas une influence minimale

Les erreurs dues au modèle peuvent provenir des données d'entrée ou du modèle direct lui-même.

La qualité des cadastres optimisés dépend de celle des cadastres *a priori* des espèces inversées. Pour Chang et al. (1997), les oscillations irréalistes obtenues lors de l'inversion des émissions de CO sont dues au fait que le cadastre *a priori* est très inhomogène. On remarquera d'ailleurs que l'on peut obtenir de meilleures concentrations avec un cadastre totalement irréaliste. Pour le NO, Elbern et al. (2000) obtiennent des résultats satisfaisants si le cadastre *a priori* n'était pas trop éloigné du régime chimique correct ; il s'avère par contre impossible d'inverser les COV individuellement. On notera de plus que le problème de la spéciation au sein des familles peut se poser : (Wang et al. 2004) suppose que la spéciation des  $\text{NO}_y$  est toujours la même, quelle que soit l'intensité des émissions de  $\text{NO}_x$  et Quélo et al. (2005) relie des mesures de monoxyde et dioxyde d'azote aux émissions de  $\text{NO}_x$ .

La méthode d'inversion choisie suppose que tout l'écart entre les mesures et les valeurs simulées est dû aux émissions à inverser. Or, **les valeurs d'autres paramètres d'entrée** du modèle direct contribuent bien sûr à cet écart, notamment les données météorologiques. Une erreur sur la direction du vent, par exemple, peut décaler le panache d'une source par rapport aux mesures ; la température joue aussi un rôle important dans la chimie photo-oxydante (Menut 2003). Les émissions inversées risquent d'être sur-corrigées pour compenser en même temps les erreurs sur les autres paramètres d'entrée, ce qui amène à des cadastres optimisés peu réalistes. (Wang et al. 2004), par exemple, déplorent un problème de reproduction de la direction du vent qui conduit à des résultats peu fiables dans certaines régions du pays étudié. Mendoza-Dominguez et Russell (2001) explorent les erreurs possibles sur deux autres types de paramètres, le flux solaire et les conditions aux limites, pour expliquer le peu d'impact de l'inversion des émissions sur les concentrations d'ozone. Elbern et Schmidt (1999) soulignent qu'il faut donc absolument éviter

ce type d'erreur et Rödenbeck et al. (2003) insistent sur l'importance d'utiliser, pour des échelles de temps de plusieurs années, des données météorologiques reproduisant la variabilité inter-annuelle.

Même si la direction et la vitesse du vent sont correctes, le **schéma de transport** qui les utilise peut avoir une précision insuffisante. À l'échelle globale, l'inversion de l'espèce utilisée comme contrainte permet de limiter les incertitudes dues aux erreurs du modèle sur certains processus. Elles ne sont cependant pas complètement éliminées. Houweling et al. (1999) estiment que la contribution des erreurs de son modèle sur le transport et la répartition spatiale des radicaux hydroxyles est «considérable». Hartley et Prinn (1993) posent la question de la qualité de la circulation dans le modèle et proposent de déduire d'inversions effectuées avec un traceur dont la chimie est bien connue l'existence et la valeur d'un éventuel biais dans la circulation. Pour contourner ce problème, Bousquet et al. (1999a) font l'hypothèse que le transport est «parfait». À petite échelle, la diffusion numérique joue un rôle crucial dans l'obtention d'un panache d'ozone de taille et de forme réaliste par exemple.

La **qualité des paramétrisations** joue un rôle déterminant. Pour la modélisation chimie-transport, les deux paramètres les plus importants sont (i) la vitesse de friction et (ii) la hauteur de couche-limite (HCL) (Menut 2003; Hanna et al. 1998; Hanna et al. 2001). La vitesse de friction  $u_*$  est utilisée directement dans le calcul du dépôt sec des espèces. À l'échelle régionale, la sensibilité des polluants à  $u_*$  n'est pas négligeable sans être la principale (Menut 2003). La hauteur de couche-limite,  $\bar{h}$ , est également un paramètre diagnostiqué à partir de certains paramètres : sa valeur, déduite des profils verticaux de vent et de température, comporte une grande incertitude. Malheureusement, elle est directement liée à l'impact des émissions surfaciques sur les concentrations de polluants : une mauvaise estimation de la hauteur de couche-limite peut amener à une trop grande concentration ou dilution des émissions. Par exemple, une surestimation des concentrations d'ozone peut être interprétée comme due à des émissions de COV ou de  $\text{NO}_x$  (selon le régime chimique (Honoré et Vautard 2000)) trop importantes ou à une faible sous-estimation de  $\bar{h}$ . Ce type de problème est particulièrement difficile à diagnostiquer puisque les paramétrisations utilisées dans les modèles sont en général les meilleures disponibles.

Il est par définition difficile de se rendre compte qu'un **processus important manque** dans un modèle. Seules des simulations à long terme et des comparaisons avec de nombreuses données permettent en général de détecter un tel manque.

Enfin, un des problèmes les plus importants est celui de la **représentativité du modèle par rapport aux mesures**. Rödenbeck et al. (2003) proposent une méthode afin d'améliorer la cohérence des comparaisons entre concentrations de dioxyde de carbone mesurées et simulées à l'échelle globale. À l'échelle régionale, au même endroit, des mesures d' $\text{O}_3$  et de  $\text{NO}_x$  n'ont pas la même représentativité spatiale, comme expliqué page 23. L'ozone est un polluant secondaire, produit loin des sources après plusieurs heures de transport. La diffusion naturelle fait donc qu'une mesure d'ozone a une représentativité qui la rend comparable à une maille de modèle à cette échelle, c'est-à-dire 3 à 6 km. Les  $\text{NO}_x$  quant à eux sont des polluants primaires qui ne sont représentatifs que de leur source, c'est-à-dire à une échelle de quelques mètres (derrière une voiture dans la circulation) à quelques centaines de mètres (près d'une autoroute) mais pas plus, ce qui est bien inférieur à la taille d'une maille.

### 2.2.3 Différence entre les échelles d'application

Les limites mathématiques de l'optimisation doivent être prises en compte dans tous les cas. Certains problèmes de stabilité peuvent se poser : pour Hartley et Prinn (1993), plus la différence entre le modèle et les observations est grande, plus l'inversion est instable et plus le biais des sources estimées est grand. De plus, il n'y a pas nécessairement unicité de la solution de la minimisation. Cependant, si la théorie mathématique est la même, les problèmes rencontrés ne sont pas les mêmes selon l'échelle spatio-temporelle considérée (Enting 2002).

Le principal obstacle technique auquel se heurtent les études en modélisation inverse est que la matrice des covariances d'erreur du modèle est inconnue. Ce problème n'est pas traité de la même façon selon l'échelle étudiée. À l'échelle globale, les auteurs essayent d'estimer les covariances grâce à des modèles de corrélation spatiaux ou basés sur des climatologies (Kasibhatla et al. 2000; Daley 1996a). À l'échelle régionale, les émissions surfaciques ne sont pas un phénomène continu, ni spatialement, ni temporellement. De plus, un flux estimé peut être faux à une heure donnée mais pas à la suivante. Il devient donc particulièrement difficile d'ajouter de l'information *a priori* dans les termes extra-diagonaux de la matrice «de background». Finalement, le plus souvent, par manque d'information, une matrice diagonale est utilisée (Elbern et Schmidt 1999; Mulholland et Seinfeld 1995).

Même avec une matrice diagonale, la valeur des termes est difficile à estimer et actualiser. Une étude récente a été menée à ce sujet mais dans le contexte de l'assimilation de données chimiques à l'échelle régionale par Hoelzemann et al. (2001). Cette méthode utilise la théorie PSAS (Physical-space Statistical Analysis System), méthode d'interpolation statistique, et permet d'élaborer un modèle de covariance inhomogène et anisotrope prenant en compte les différences entre milieux urbain et rural et la densité du réseau d'observations. Cette méthode, basée sur des informations statistiques, est tout à fait correcte pour des simulations à long terme (par exemple des scénarios de variation d'émissions d'année en année) mais ne peut être utilisée à l'échelle régionale et horaire à cause de la forte variabilité des sources et des puits.

L'échelle de temps des phénomènes considérés et notamment la chimie (Enting 2002) amène de plus des problèmes particuliers à l'échelle régionale. Pour Mulholland et Seinfeld (1995), le décalage entre les émissions et l'impact sur les mesures peut être particulièrement gênant et difficile à prendre en compte. Chang et al. (1996) notent que les corrections apportées aux émissions à une heure donnée sur-compensent souvent les mauvaises estimations des heures précédentes. La première heure d'inversion est corrigée de façon à compenser de mauvaises conditions initiales et des solutions comportant des émissions négatives sont obtenues pour certaines heures. Ces résultats irréalistes sont interprétés comme indiquant soit une sur-compensation due à des surestimations aux heures précédentes soit un mécanisme manquant dans le modèle.

Finalement, toutes les études effectuées indiquent que les méthodes inverses sont très prometteuses mais que de nombreuses limites demeurent.

## 2.3 Élaboration de notre méthode d'inversion

Le fait que les émissions matinales de certaines espèces (les oxydes d'azote) influencent le pic d'ozone de l'après-midi nécessite de disposer d'une méthode permettant de prendre en compte la trajectoire du modèle. Or, originellement, les méthodes basées sur le filtre de Kalman ne pouvaient traiter que de trois dimensions, sans prendre en compte le temps. L'approche adjointe (voir page 31) a donc été choisie pour cette étude parce qu'elle permet de prendre en compte les liens non-linéaires entre émissions et concentrations le long de trajectoires à des échelles de temps égales au temps de vie chimique des espèces étudiées.

### 2.3.1 Le vecteur d'état

Le vecteur d'état contient en principe les paramètres à inverser. Comme ceux-ci pourront être de natures différentes (flux d'émissions ou paramètres météorologiques) et couvrir plusieurs ordres de grandeur (pour les flux d'émissions), on utilisera en pratique un vecteur d'état  $x^*(t)$  contenant les **coefficients correctifs** à appliquer aux paramètres correspondant. Le vecteur  $x(t)$  est reconstitué en appliquant  $x^*(t)$  aux paramètres *a priori*.

Pour prendre en compte la dimension temporelle du problème, le vecteur d'état contient les paramètres de plusieurs heures comme autant de composantes différentes. En effet, les espèces émises n'ont pas une influence immédiate sur les espèces mesurées. Le pic d'ozone de l'après-midi par exemple est principalement sensible aux émissions de NO<sub>2</sub> du pic de trafic du matin. Il apparaît donc nécessaire d'effectuer l'inversion dans des **fenêtres temporelles** de taille pertinente par rapport à la physique et à la chimie de la pollution photo-oxydante, par exemple, de 9 heures à 15 heures pour l'inversion du NO<sub>2</sub> avec des mesures d'ozone.

La fonction coût prend donc en compte les mesures disponibles pendant toute la «fenêtre» et une seule exécution de l'adjoint (en remontant le temps pour toute la période) permet d'obtenir toutes les composantes du gradient. Il faut remarquer que l'utilisation de fenêtres temporelles augmente rapidement la dimension du problème ; leur durée ne peut donc pas dépasser quelques heures et dépend des liens entre espèces mesurées et inversées.

### 2.3.2 La matrice $R$

Un des objectifs de ce travail est d'élaborer une méthode utilisant les **mesures de routine des réseaux**. Or, pour un réseau de mesures, les covariances des erreurs d'observation c'est-à-dire les termes extra-diagonaux de  $R$ , sont supposés nuls car tous les instruments sont indépendants. Il reste donc à estimer les termes diagonaux de  $R = \text{diag}(w_1, \dots, w_N)$  (avec  $N$  le nombre de contraintes disponibles).

Pour prendre en compte la qualité des contraintes c'est-à-dire la confiance qu'on leur accorde, les poids des termes de la fonction coût comportant une différence entre concentrations simulée et mesurée sont proportionnels à la confiance que l'on accorde à la contrainte. Le poids  $w_{ii}$  doit donc être proportionnel à la confiance accordée à  $y_{obs}(i)$ . On pourrait utiliser  $w_i \propto \frac{1}{var(i)}$  avec  $var(i)$  la variance associée à  $y_{obs}(i)$  et  $I(i)$  l'incertitude correspondant. Mais si l'on utilise la variance absolue, de petites concentrations auront une petite variance et donc un grand poids dans l'optimisation alors que les concentrations élevées auront une variance absolue beaucoup plus grande et donc une influence moindre



dans l'optimisation. Or, c'est notamment la reproduction correcte des pics de concentrations qui est recherchée. Au lieu d'utiliser directement la variance des contraintes, on utilise donc l'incertitude relative  $I_{rel}$  transformée en «variance relative» :  $var(i) = I(i)^2$  et  $I_{rel}(i) = I(i)/val(i)$  donnent  $var_{rel}(i) = I_{rel}(i)^2 = (I(i)/val(i))^2$ . Pour plus de facilité de lecture, tout est exprimé en pourcentage :  $I_{rel}(i)(pourcent) = I(i)/val(i) \times 100$  d'où  $var_{rel}(i)(pourcent) = var_{rel}(i) \times 10000$ .

### 2.3.3 La matrice $B$

Dans le cas de notre étude, aucune information sur l'incertitude du cadastre *a priori* n'est disponible. Celle-ci pourrait être de 35% (Kühlwein et Friedrich 2000) pour les espèces anthropiques et d'un facteur 2 à 3 pour les espèces biogéniques (Beekmann et Derognat 2003; Sillman et al. 2003). Finalement, l'espace des émissions ne peut pas être enrichi par des informations *a priori* permettant de construire une matrice  $B$  **pour les émissions**.

Dans le cas de l'inversion de paramètres météorologiques, une matrice «de background» au moins diagonale devra être élaborée. En effet, la météorologie étant calculée «off-line», il est nécessaire d'éviter que les valeurs optimisées de certains paramètres ne s'éloignent trop des valeurs *a priori*, sous peine d'obtenir des résultats incohérents. Dans le cas de la hauteur de couche-limite par exemple, la valeur optimisée ne doit pas s'éloigner de la valeur *a priori* au-delà de l'incertitude de la fonction de diagnostic qui dépend de la température et de la pression fournies «off-line». Dans le cas contraire, la hauteur de couche-limite optimisée n'est plus en accord avec les profils verticaux de température et de pression.

## 2.4 Conclusion

Un état de l'art des méthodes inverses a permis de constater que celles-ci fonctionnent depuis longtemps de façon satisfaisante à l'échelle globale. Les principales limitations sont dues aux incertitudes et aux erreurs des mesures, du modèle et des données d'entrée. À l'échelle régionale, cependant, peu d'études sont disponibles et les problèmes rencontrés ne sont pas tout à fait les mêmes qu'à l'échelle globale.

La méthode de l'adjoint a été choisie pour notre travail parce qu'elle permet de prendre en compte facilement la trajectoire du modèle. La fonction coût élaborée est d'une forme classique : elle comporte (i) un terme constitué par les différences entre contraintes et prévisions avec une matrice de covariances d'erreur d'observation diagonale dont les termes sont calculés simplement à partir des variances associées aux contraintes (ii) un terme constitué par la différence entre les valeurs des paramètres *a priori* et optimisés avec une matrice de covariance d'erreur du modèle à élaborer selon les paramètres inversés.

Parmi les limitations rencontrées, les plus préoccupantes sont la quantité insuffisante de contraintes par rapport à la dimension du problème, la pertinence des informations disponibles dans ces contraintes par rapport au problème posé et enfin la qualité de ces informations. Pour contourner ces difficultés, deux stratégies sont envisageables : d'une part, augmenter le nombre de contraintes et leur «adéquation» par rapport au problème à inverser et d'autre part, diminuer la dimension du problème. Elles ont été mises en œuvre simultanément en utilisant les informations contenues dans l'espace des concentrations c'est-à-dire des prévisions et dans celui des paramètres pour les émissions.

# Chapitre 3

## Cadre de modélisation

Pour procéder à l'inversion des émissions à l'échelle régionale, il est nécessaire de disposer d'un modèle permettant :

- de prendre en compte tous les phénomènes physiques et chimiques pertinents à cette échelle,
- d'effectuer des simulations à une résolution spatiale inférieure à la dizaine de kilomètres et à une résolution temporelle de l'ordre de la minute,
- de prendre en compte des conditions aux limites ayant une résolution temporelle de l'ordre de l'heure (et non des valeurs climatologiques).

De plus, ce modèle doit être «adjointisable» c'est-à-dire qu'il doit être construit de telle sorte qu'il est possible d'obtenir un adjoint fonctionnel. Enfin, le modèle direct doit avoir été validé pour les régions dans lesquelles les études vont être menées, notamment l'Île-de-France.

Ces différentes exigences nous ont conduit à choisir le modèle CHIMERE.

### 3.1 Le modèle CHIMERE

#### 3.1.1 Principe général du modèle

CHIMERE est un modèle eulérien tridimensionnel : il simule sur un maillage cartésien régulier des champs de concentrations d'espèces chimiques d'intérêt pour l'étude de la pollution atmosphérique dans la basse troposphère et notamment la couche-limite atmosphérique (Vautard et al. 2001; Schmidt et Martin 2003).

Comme tout modèle de chimie-transport, CHIMERE intègre le système différentiel représentant l'évolution d'espèces chimiques à partir de leurs sources et leurs puits :

$$\frac{\partial C_i}{\partial t} + \nabla \cdot (V.C_i) = \nabla \cdot (K.\nabla.C_i) + PL_i + E_i - D_i \quad (3.1)$$

où  $C_i$  est la concentration de l'espèce chimique  $i$ ,  $K$  le coefficient de diffusivité verticale,  $PL_i$  le bilan net de produit et perte sur cette espèce,  $E_i$  les émissions surfaciques et  $D_i$  le dépôt sec.

À l'origine, CHIMERE a été développé dans un but de prévision déterministe des concentrations d'ozone en région parisienne (Vautard et al. 1997), la prévision des pics d'ozone étant un de ses principaux objectifs. La première version du modèle était un modèle de

boîtes, représentant les départements d'Ile-de-France (Menut et al. 2000). Depuis, toute la structure du modèle a évolué, pour en faire un outil précis à même de simuler, en peu de temps informatique, de longues périodes et sur de grands domaines.

### 3.1.2 Les paramètres d'entrée

#### 3.1.2.1 La géométrie des domaines

À partir d'un unique code, CHIMERE peut fonctionner à deux échelles : (i) à l'échelle continentale, pour un domaine qui couvre l'Europe de l'Ouest avec une résolution de  $0.5^\circ$ , comme représenté sur la Figure 3.1 par les points indiquant les nœuds du maillage et (ii) à l'échelle régionale pour divers domaines dont la position est indiquée sur la Figure 3.1 par les rectangles, avec généralement des résolutions de 3 à 6 km.

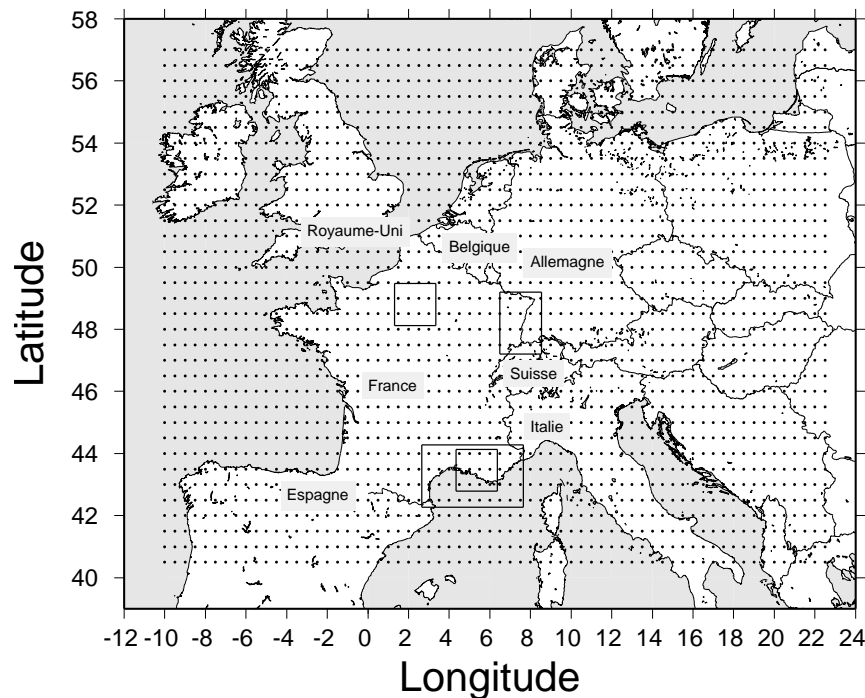


FIG. 3.1 – Cartes des domaines de simulation CHIMERE

En mode prévision, pour garantir la rapidité du calcul, le modèle utilise huit niveaux verticaux, allant de la surface à 750 hPa (à peu près 3000 mètres d'altitude), leur épaisseur dépendant de la météorologie. Les caractéristiques du sol et le relief des différents domaines sont interpolés à partir des données de la NOAA (National Oceanic and Atmospheric Administration, [www.noaa.gov](http://www.noaa.gov)).

#### 3.1.2.2 La météorologie

CHIMERE est un modèle «off-line» et doit donc être alimenté par des données météorologiques précalculées. En mode prévision, pour le domaine continental, les données utilisées sont celles délivrées quotidiennement par le Centre Européen pour les Prévisions

Météorologiques à Moyen Terme (CEPMMT ou ECMWF en anglais) avec une résolution d'un demi degré. Pour les domaines régionaux, ces mêmes données sont interpolées, de manière à fournir des champs météorologiques à la résolution voulue.

### 3.1.2.3 Les conditions aux limites et les conditions initiales

Les conditions aux limites ne concernent que les espèces chimiques. La version continentale utilise des champs de concentrations climatologiques calculés avec le modèle de chimie-transport global MOZART (Brasseur et al. 1998). L'intérêt de l'approche à deux échelles est de permettre le «nesting» : les résultats des simulations à l'échelle continentale sont utilisées comme conditions aux limites pour les domaines régionaux. Une simulation régionale inclut donc de manière systématique une première simulation continentale sur la même période.

Les conditions initiales ont un impact fort sur les concentrations calculées au début d'une simulation. À l'échelle continentale, une simulation doit prendre en compte le transport des masses d'air par le vent sur plusieurs centaines de kilomètres. Le temps nécessaire pour que les conditions initiales n'aient plus d'influence sur les concentrations simulées est donc de plusieurs jours. À l'échelle régionale, les conditions initiales n'ont pas d'impact fort sur les concentrations au-delà d'un cycle diurne. Selon l'échelle, il est donc nécessaire d'effectuer un ou quelques jours de simulation non exploitables avant la période d'intérêt.

### 3.1.2.4 Les émissions

En version prévision, il faut fournir au modèle les émissions de seize espèces-modèle sous forme de flux surfaciques exprimés en molécules par centimètre carré par seconde ( $\text{moléc.cm}^{-2}.\text{s}^{-1}$ ).

**À l'échelle continentale** Les émissions anthropiques sont issues de la base de données annuelles EMEP pour 1998 sous la forme de masses totales annuelles de  $\text{NO}_x$ ,  $\text{SO}_2$ , CO et de composés organiques volatils (COV) non méthaniques (éthane, n-butane, éthène, propène, o-xylène, formaldéhyde, acétaldéhyde et méthyle, éthyle et cétone). Il est nécessaire de transformer ces données brutes en profils temporels de résolution horaire, dépendant du mois, du jour (semaine, jour férié) et de l'heure, pour chacune des espèces-modèle.

La ventilation temporelle des émissions s'effectue grâce à des coefficients à appliquer aux masses annuelles pour les moduler selon le mois, le jour de la semaine et l'heure de la journée. Ces profils temporels sont construits à partir des informations contenues dans la base de données GENEMIS (GENeration and evaluation of EMISsion data) (Society 1994). La modulation mensuelle prend en compte des variations saisonnières comme celles des émissions dues au chauffage. Les profils horaires prennent en compte trois types de jour correspondant aux jours de la semaine (du lundi au vendredi), au samedi et au dimanche ou jour férié. Les émissions de tous les lundis ont donc un profil temporel identique à celui de tous les mardis.

Une fois les masses annuelles réparties dans le temps, il faut effectuer la spéciation des espèces en espèces-modèles. Après avoir été regroupés en trente-deux classes selon leur

réactivité, les COV non méthaniques sont agrégés en 10 espèces-modèles suivant les recommandations de Middleton et al. (1990).

À ces émissions anthropiques, sont ajoutées des émissions biogéniques d'isoprène, de terpènes et de NO, calculées à partir des données météorologiques et des caractéristiques du terrain (Guenther 1997).

**À l'échelle régionale** Les émissions employées pour l'Île-de-France dans le cadre de cette étude ont été élaborées par AIRPARIF pour la campagne ESQUIF sur la base du mois de Juillet 1998 à l'aide des modules «émissions» de SIMPAR (Système Informatique de Modélisation de la Pollution Atmosphérique à l'échelle Régionale). Ce qui suit résume les explications détaillées fournies par Vautard et al. (2003) et dans le rapport ESQUIF (2001).

Le cadastre élaboré indique les émissions de monoxyde de carbone, dioxyde de carbone, dioxyde de soufre, oxydes d'azote, COV non méthaniques et méthane pour trois types de jours correspondant aux jours de semaine, au samedi et au dimanche (ou jour férié) à une résolution spatiale de 3 km (sur le domaine de 150 km × 150 km représentant l'Île-de-France) avec un pas horaire d'une heure.

Les masses totales annuelles sont calculées en prenant en compte les onze secteurs d'activité définis par la nomenclature SNAP (Selected Nomenclature for sources of Air Pollution) dont les intitulés sont les suivants :

- 1 - combustion dans les industries de l'énergie et de la transformation de l'énergie
- 2 - combustion hors industrie
- 3 - combustion dans l'industrie manufacturière
- 4 - procédés de fabrication
- 5 - extraction et distribution de combustibles fossiles
- 6 - utilisation de solvants et autres produits
- 7 - transport routier
- 8 - autres sources mobiles et machineries
- 9 - traitement et élimination des déchets
- 10 - agriculture et sylviculture et aquaculture
- 11 - autres sources et puits

En pratique, les sources sont divisées en trois catégories pour chacune desquelles un mode de calcul adapté est appliqué :

- (i) les grandes sources ponctuelles telles que cheminées industrielles, centrales thermiques, incinérateurs, soit en tout deux cent trois en Île-de-France. Les émissions sont calculées à partir des données collectées par la DRIRE (Direction Régionale de l'Industrie, de la Recherche et de l'Environnement) d'Île-de-France en 1998. Les profils temporels permettant de passer des masses totales annuelles à la résolution horaire sont tirés du programme GENEMIS (comme à l'échelle continentale) et des déclarations des industries.
- (ii) les sources linéiques (grands axes routiers, autoroutes). Elles sont calculées en combinant quatre types d'information :

- les données sur le réseau routier et le trafic, fournies par la DREIF (Direction Régionale de l'Équipement en Île-de-France) et qui concernent les types de voies (voies urbaines, routes et autoroutes), leur débit, la vitesse du trafic, le pourcentage de véhicules roulant «à froid» pour les trente cinq mille brins du réseau.
- la composition du trafic c'est-à-dire les pourcentages de chaque catégorie de véhicules (véhicules particuliers, deux roues, véhicules utilitaires légers, poids lourds, bus et cars) circulant sur les trois types de voies selon l'heure de la journée.
- les facteurs d'émissions pour ces véhicules, obtenus d'après la méthode COPERT III (Computer Programme to Calculate Emissions from Road Transport) (Ntziachristos et al. 2000)
- la température extérieure, qui influence les émissions à froid et les pertes par évaporation.

Les profils temporels diurnes sont calculés par le modèle de trafic mais une modulation, obtenue à partir des mesures d'une station «trafic» du réseau AIRPARIF, est appliquée pour différencier les trois types de jours.

- (iii) les sources surfaciques c'est-à-dire celles qui ne font pas partie des deux premières catégories comme le chauffage domestique, le trafic diffus, les petites sources industrielles. Les données du CITEPA (Centre Interprofessionnel Technique d'Études de la Pollution Atmosphérique) pour 1994, fournies pour des arrondissements ou départements, sont combinées à l'occupation des sols indiquée par l'IFEN (Institut Français de l'ENVironnement) afin d'obtenir la résolution spatiale voulue. Les profils temporels proviennent également de GENEMIS.

La spéciation des COV en espèces-modèle est effectuée pour chaque secteur SNAP d'après des études diverses (Theloke et al. 2000; Theloke et al. 2001; Schmitz et al. 2000; Rudd et Marlowe 1998; Hassel et al. 2000). Les oxydes d'azote sont répartis en 10% de dioxyde d'azote et 90% de monoxyde d'azote.

Dans le cadastre obtenu, le trafic routier est la principale source de  $\text{NO}_x$ , avec 53% des émissions totales. Les solvants et le trafic sont les principales sources de COV (36 et 35% respectivement).

### 3.1.3 Principes généraux des schémas et paramétrisations

**La chimie** CHIMERE peut travailler avec divers mécanismes chimiques. En prévision, une version réduite du mécanisme chimique MELCHIOR (Modèle d'Étude Lagrangienne de la CHimie de l'Ozone à l'échelle Régionale) est utilisée. Elle comporte 40 espèces et 100 réactions (80 espèces chimiques et 200 réactions pour le mécanisme complet) (Latuati 1997). Le rayonnement est pris en compte grâce à des taux de photolyse tabulés, variables par ciel clair (Madronich et Flocke 1998). Un modèle statistique prenant en compte la nébulosité et le profil d'humidité dans les couches basses permet de paramétrer l'atténuation du rayonnement due aux nuages (ESQUIF 2001).

**La turbulence et le dépôt** Des paramètres turbulents doivent être diagnostiqués à partir des grandeurs météorologiques principales (Louis 1979; Louis et al. 1982). On doit donc estimer le flux de chaleur sensible surfacique  $Q_0$ , la vitesse de rugosité  $u_*$ , la hauteur de la couche limite atmosphérique  $\bar{h}$ , le profil de diffusion turbulente  $K_z$ . Ces différentes paramétrisations sont décrites en détail dans (Vautard et al. 2001; Menut 2003) par exemple.

Le dépôt sec est paramétré suivant un modèle de résistances (voir page 21). La résistance aérodynamique est inversement proportionnelle au coefficient de traînée (calculé par la méthode de Louis (1979)) et au module du vent dans la couche inférieure du modèle (Wesely et Hicks 1977). La résistance de la couche laminaire est approchée par la formule de Hicks et al. (1987) avec une vitesse de friction estimée selon Louis (1979). Enfin, la résistance de surface dépend de la solubilité et du pouvoir oxydant de l'espèce chimique, qui interviennent dans la paramétrisations de processus en surface des végétaux (Erisman et al. 1994), ou prend en compte le type de sol selon la base de données de van de Velde et al. (1994).

**Le transport** Le transport horizontal peut être traité grâce au schéma PPM (Parabolic Piecewise Method) d'ordre trois de Colella et Woodward (1984) ou grâce au schéma dit «Van Leer» (Van Leer 1994) dont le code est reproduit en Annexe C.

Le transport vertical est déterminé pour assurer la conservation de la masse : il est calculé de façon à compenser la convergence ou la divergence des flux horizontaux.

**Résolution numérique** Une caractéristique importante de CHIMERE par rapport à d'autres modèles est qu'il n'intègre pas le système 3.1 selon l'approche classique de séparation des opérateurs (Figure 3.2). Le modèle intègre au contraire globalement tous les flux pour une même espèce. Cette intégration est réalisée avec le solveur de systèmes différentiels raides Two-Step (Lanser et Verwer 1998).

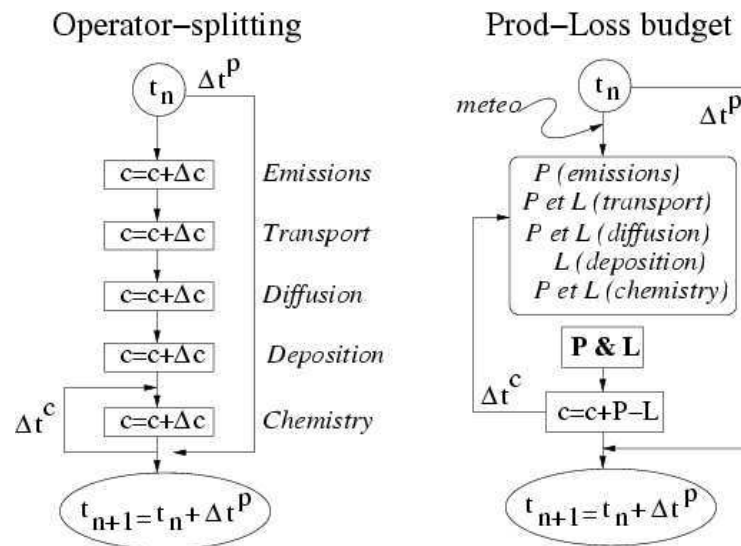


FIG. 3.2 – Différence entre la méthode de séparation des opérateurs («operator splitting») et la méthode utilisée dans CHIMERE («Prod-Loss budget»).

CHIMERE est utilisé actuellement en prévision par AIRPARIF à Paris mais aussi à Lyon, à Toulouse, à Marseille et en Alsace et à l'échelle européenne dans le système PREV' AIR (<http://prevair.ineris.fr/fr/introduction.php>). En recherche, CHIMERE est utilisé pour des études de cas et de scénarios : parmi les projets en cours, on peut citer notamment la campagne ESCOMPTE, menée à l'été 2001 dans la région de Marseille et

comportant un exercice de modélisation (disponible à [medias.obs-mip.fr/escompte/exercice/HTML/exe.html](http://medias.obs-mip.fr/escompte/exercice/HTML/exe.html)). L'inter-comparaison de nombreux modèles de chimie-transport, dont CHIMERE, fait l'objet du projet City-Delta (disponible sur <http://rea.ei.jrc.it/netshare/thunis/citydelta/>). De nombreux développements de CHIMERE sont en cours, notamment la prise en compte des poussières désertiques (CHIMERE-DUST) et des particules (Bessagnet et al. 2004).

## 3.2 L'adjoint

### 3.2.1 Principe

Alors que le modèle direct peut calculer l'impact de la modification d'un paramètre sur les prévisions, le modèle adjoint calcule pour une prévision l'impact de la variation de tous les paramètres, comme schématisé sur la Figure 3.3. Comme le problème n'est pas linéaire, l'impact calculé n'est valable que pour une variation infinitésimale du paramètre.

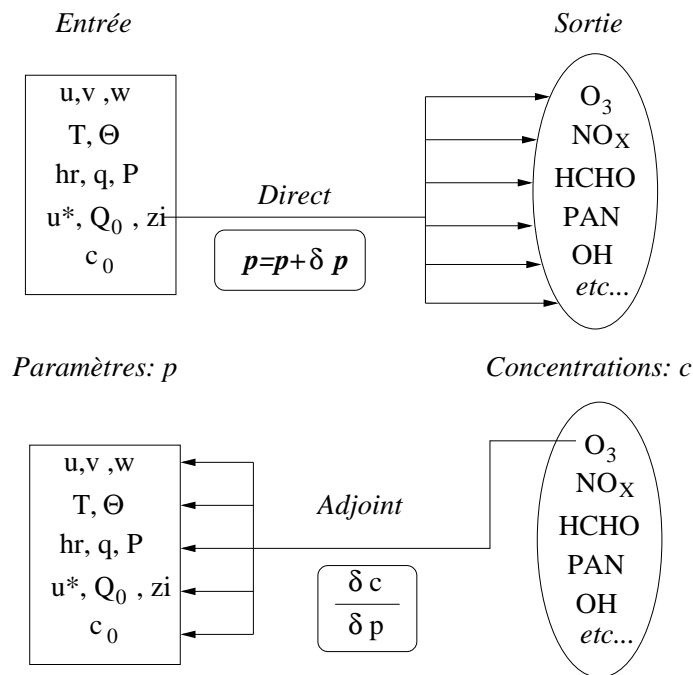


FIG. 3.3 – Principe de l'adjoint d'un modèle de chimie-transport.

D'un point de vue plus mathématique, l'adjoint a pour but de calculer le gradient d'une fonction scalaire des paramètres. Le modèle adjoint est donc le gradient du modèle direct.

### 3.2.2 Développement et validation

Les équations adjointes sont obtenues à partir de modèle linéaire tangent qui est la linéarisation du modèle direct (non-linéaire) autour des variables directes d'entrée. L'opérateur d'adjonction est défini de telle sorte qu'il est équivalent à une transposition si on utilise le produit scalaire canonique. À partir du modèle linéaire tangent, une simple transposition



des lignes en colonne (ainsi que l'inversion de l'ordre d'actualisation des variables) permet d'obtenir les équations adjointes, les variables directes d'entrée devenant les variables adjointes de sortie.

En pratique, pour construire l'adjoint d'un modèle, il faut donc d'abord écrire le modèle linéaire-tangent puis l'adjointiser. Il existe des programmes qui effectuent cette tâche automatiquement. L'adjoint de CHIMERE version 200211K a cependant été réalisé «à la main» afin d'obtenir une programmation permettant d'optimiser le temps de calcul. Une intégration adjointe est en effet environ trois fois plus longue qu'une intégration directe. À chaque version du modèle direct correspond donc un adjoint, qui doit être mis à jour lorsque le modèle direct évolue. Dans le cadre de ce travail de thèse, la version de l'adjoint correspondant à la version 200211K du modèle direct CHIMERE a été complétée et actualisée. Le code adjoint du schéma de transport «Van Leer», qui a été élaboré au cours de ce travail est reproduit en Annexe C à titre d'exemple.

La validation de l'adjoint est une étape délicate, décrite en Annexe C.

### 3.3 L'optimiseur

Le minimiseur utilisé est N1QN3 (Gilbert et Lemaréchal 1989), dans la version décrite plus récemment par Gilbert et Lemaréchal (1995). Il utilise une méthode de type quasi-Newton à mémoire limitée pour résoudre les problèmes de minimisation sans contrainte. Son principe consiste en la détermination à chaque itération d'une direction de descente et d'un pas adapté le long de cette direction. Celle-ci est calculée en utilisant une approximation du hessien et le gradient de la fonction coût au point courant. La longueur du pas est obtenue par une procédure de recherche linéaire et doit satisfaire les conditions de Wolfe. Les itérations s'arrêtent lorsque la précision choisie pour le critère d'arrêt est atteinte. Ce critère est le rapport entre les normes du gradient au point courant et au point de départ. La convergence est alors considérée comme atteinte au point courant. Des nombres maximaux d'itérations et de simulations sont choisis selon le temps de calcul considéré comme raisonnable. Si l'optimisation ne converge pas avant qu'un de ces deux nombres ne soit atteint, elle s'arrête au point atteint.

### 3.4 L'inverse

#### 3.4.1 Principe : imbrication des modules de calcul

L'inverse ne représente en principe qu'un gestionnaire permettant de calculer le coût et son gradient puis de procéder à leur minimisation. Il s'agit donc de faire appel au modèle direct, à l'adjoint et au minimiseur.

La Figure 3.4 indique les différentes entrées et étapes d'une simulation inverse. La procédure d'inversion comprend cinq étapes :

Étape 1 : le modèle est initialisé avec les paramètres *a priori*  $x_f$  et les concentrations  $y_{sim}$  sont calculées

Étape 2 : une première valeur du coût  $J(x(t_0)) = J(x_f)$  est évaluée avec les contraintes  $y_{obs}$  disponibles

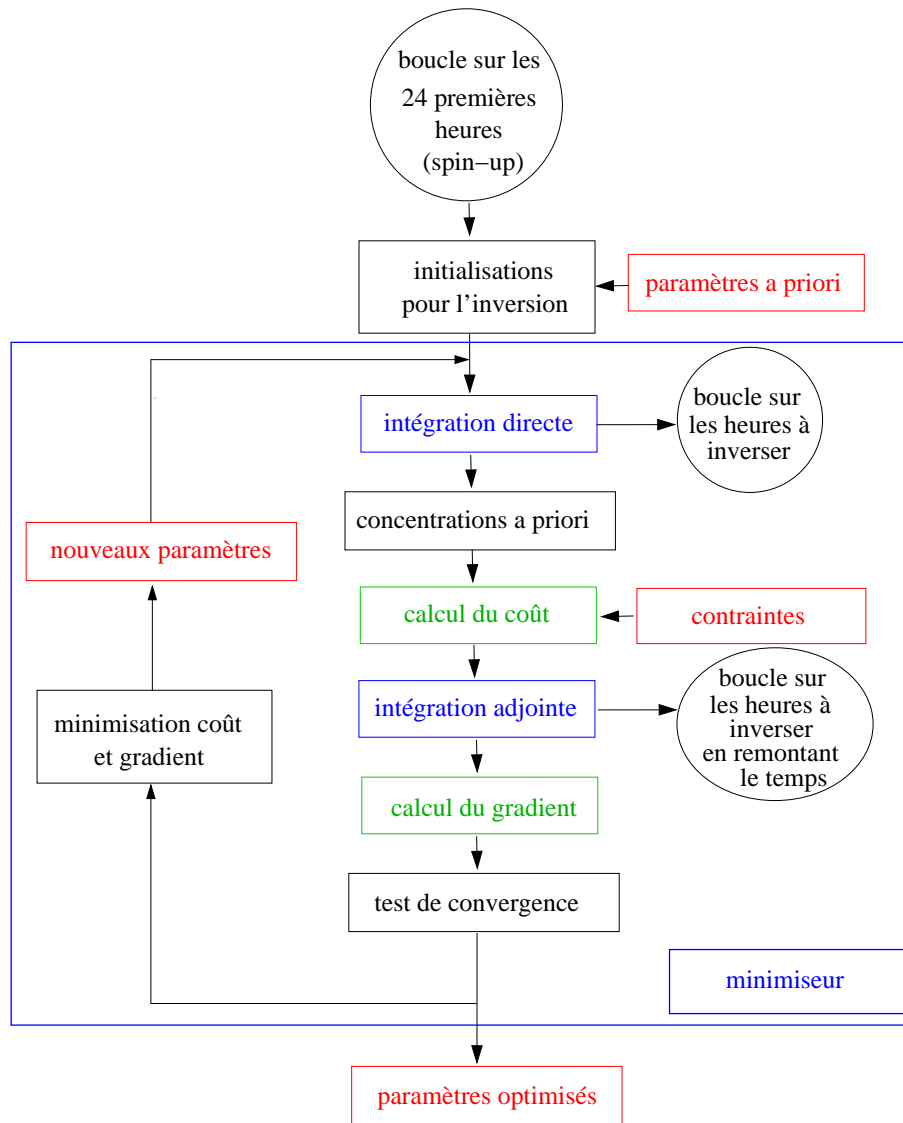


FIG. 3.4 – Principe de l'inversion d'une simulation.

Étape 3 : une simulation adjointe est faite et les composantes du gradient de la fonction coût  $\frac{\partial J}{\partial x}(x_f)$  sont calculées

Étape 4 : le minimiseur fournit le meilleur ajustement  $x(t_1)$  de  $x$  pour minimiser  $J$  et  $\partial J / \partial x$

Étape 5 : le modèle est ré-initialisé avec les nouveaux paramètres  $x(t_1)$

Les itérations se poursuivent jusqu'à ce que le critère d'arrêt soit atteint.

L'inverse basé sur la version V200211K de CHIMERE a été développé dans le cadre de ce travail de thèse. Son organigramme et l'algorithme tel qu'il a été programmé constituent l'Annexe D.

### 3.4.2 Le critère d'arrêt

Le critère de convergence sur la norme du gradient est noté *epsg*. L'optimisation s'arrête quand le nouveau rapport  $\epsilon_k$  calculé est inférieur à *epsg* choisi au départ :

$$\epsilon_k = \frac{\|g_k\|}{\|g_1\|} < epsg.$$

On veut s'arrêter quand  $\|g_k\|$  est presque nulle (en supposant qu'elle peut l'être).

$$\|g_k\| = \sqrt{\sum_{n=1}^{dim} \frac{\partial J}{\partial X_n}^2}$$

$$\frac{\partial J}{\partial X_n} = \lim_{h \rightarrow 0} \frac{J(\dots, X_n + h, \dots) - J(\dots, X_n, \dots)}{h}$$

Mais comme le vecteur  $x$  contient des coefficients correctifs, on ne cherche pas à faire tendre  $h$  vers 0 pour ne pas tenir compte de corrections inférieures à 1% par exemple. On cherche plutôt une situation où  $\frac{\partial J}{\partial X_n}$  est quasiment nulle avec un  $h = 10^{-2}$  fini. Il ne reste plus alors qu'à estimer une valeur pour laquelle la différence  $\Delta J = J(\dots, X_n + h, \dots) - J(\dots, X_n, \dots)$  est à peu près nulle.  $\Delta J$  est de la forme :

$$\Delta J = \sum_{i=1}^N w_i [(y_{obs}(i) - y_{sim}(i, X_n + h))^2 - (y_{obs}(i) - y_{sim}(i, X_n))^2]$$

La différence entre deux termes en ppb est non significative quand elle est strictement inférieure à  $10^{-2}$ .  $\Delta J$  est donc non significative quand elle est strictement inférieure à  $10^{-4} \times \sum_{i=1}^N w_i$ . Pour une meilleure approximation, on considère que  $\Delta J$  «nulle» quand elle est égale à  $10^{-5} \times \sum_{i=1}^N w_i$

Finalement, on obtient une valeur  $\|g_k\|$  satisfaisante quand on a :

$$\begin{aligned} \|g_k\| &\approx \sqrt{dim \times \left( \frac{10^{-5} \times \sum_{i=1}^N w_i}{10^{-2}} \right)^2} \\ &\approx \sqrt{dim} \times \frac{10^{-5} \times \left\| \sum_{i=1}^N w_i \right\|}{10^{-2}} \\ &\approx \sqrt{dim} \times 10^{-3} \times \sum_{i=1}^N w_i \end{aligned}$$

On peut donc choisir

$$epsg = \frac{\sqrt{dim} \times 10^{-3} \times \sum_{i=1}^N w_i}{\|g_1\|}.$$

## 3.5 Un outil d'étude méthodologique : les cas académiques

### 3.5.1 Principe

L'étude de cas académiques est utilisée pour la validation de méthode comme l'assimilation de données (Elbern et Schmidt 1999; Cirpka et Kitanidis 2001). Elle permet de valider la méthodologie employée tout en déterminant ses limitations.

L'élaboration d'un cas académique permet en effet de disposer d'une «réalité» à laquelle il suffit de comparer les résultats de l'optimisation pour déterminer le degré de précision de la méthode. La démarche d'inversion d'un cas académique est résumée sur la Figure 3.5. La simulation directe effectuée avec le cadastre «réel» (qui donne donc les valeurs de concentrations considérées comme «réelles» pour tous les polluants dans toutes les mailles du modèle) et le choix des stations de mesure (emplacement et espèces mesurées) permettent de générer les mesures, plus ou moins entachées d'erreurs selon l'incertitude choisie. Il faut noter que les problèmes dus à la diffusion numérique sont en partie évités puisque mesures et «réalité» sont générées avec le même maillage. Les paramètres *a priori* sont créés en perturbant les paramètres «réels». Dans le cas du cadastre d'émission indiqué sur la Figure 3.5, les émissions de certaines espèces sont multipliées ou divisées. La simulation inverse utilise le cadastre *a priori* (qui fournit donc les concentrations simulées *a priori*) et les mesures pour produire un cadastre optimisé. La comparaison entre cadastre «réel» et cadastre optimisé est possible, contrairement à ce qui se passe dans la réalité.

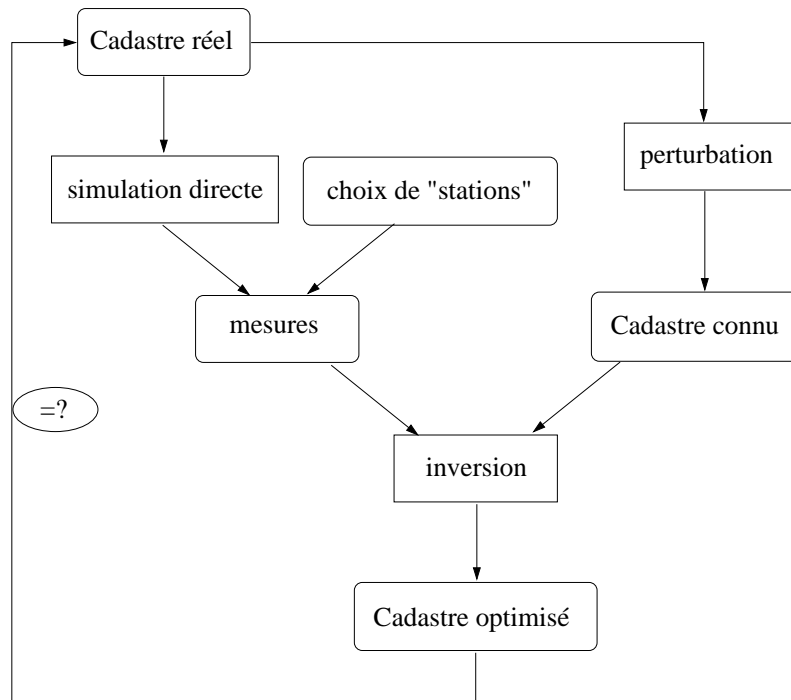


FIG. 3.5 – Principe d'élaboration d'un cas académique

### 3.5.2 Du simple au quasi-réel

Le principe général décrit à la Section précédente permet d'élaborer différents types de cas académiques.

Les cas académiques formels, comme ceux de la Section suivante, utilisent des fichiers de paramètres (météorologie et émissions) élaborés spécialement représentant des situations extrêmement simples. Le maillage du domaine compte moins de boîtes afin de raccourcir le temps de calcul et de simplifier la visualisation des résultats.

Les cas académiques-réels sont basés sur de véritables simulations, telles qu'elles ont déjà été effectuées, par exemple en prévision. Elles constituent alors la «réalité» et les valeurs de certains paramètres, par exemple quelques flux d'émission, sont perturbés pour générer le cas *a priori*.

Dans tous les cas, seuls les paramètres sont modifiés, notamment le cadastre d'émissions, la météorologie ou le domaine ; les mécanismes internes au modèle sont inchangés, par exemple la chimie et les paramétrisations de transport.

## 3.6 Validation de la gestion numérique

Afin de valider la gestion numérique de l'inverse, des cas académiques formels ont été élaborés. Le nombre de mailles réduit assure des calculs rapides et la simplicité des problèmes permet de mettre en évidence les principaux aspects du processus d'inversion.

### 3.6.1 Un modèle simple de cas académiques

#### 3.6.1.1 Maillage et météorologie

Le domaine représente l'Île-de-France, de 1,3 à 3,3°E et de 48,1 à 49,4°N. Le maillage est constitué de  $10 \times 10$  mailles pour cinq niveaux verticaux dont l'épaisseur ne varie pas dans le temps. L'épaisseur du premier niveau est fixée à 20 mètres, les niveaux supérieurs sont de plus en plus épais et englobent jusqu'à 300 mètres au-dessus de la hauteur de couche-limite maximale (Figure 3.7).

Les conditions météorologiques reproduisent celles d'un épisode de pollution estival en Île-de-France : températures élevées, des vitesses de vent faibles et aucun nuage. Le vent et la température sont uniformes dans le domaine.

La direction du vent est fixée à 45° c'est-à-dire soufflant du Nord-Est à une vitesse en surface de  $4 \text{ m.s}^{-1}$ . Cette vitesse correspond à une valeur réaliste en Île-de-France (voir par exemple la POI 5 d'ESQUIF, Annexe B) assez élevée pour permettre une bonne visualisation des résultats dans le maillage choisi. Le profil vertical du module du vent est déterminé à partir de cette vitesse de surface et une loi logarithmique dans la couche-limite. Au-dessus de la couche limite, le module du vent est constant.

La valeur moyenne de la température à deux mètres pour la journée est fixée à 23°C. La température à 2 mètres suit un profil temporel sinusoïdal (Figure 3.6) qui représente un cycle diurne typique avec un maximum à midi et un minimum une heure avant le lever du Soleil. Un profil vertical réaliste est obtenu en appliquant une décroissance linéaire de 6K par kilomètre avec une inversion de 2K à la hauteur de la couche limite.

La hauteur de couche-limite minimale est fixée à 300 mètres, la hauteur maximale est fixée à 2200 mètres. Le Soleil se levant à 5h et se couchant à 19h (il est au zénith à 12h), le maximum est atteint à 15h (trois heures après le passage du Soleil au zénith). L'évolution de la hauteur de couche-limite est sinusoïdale pendant les deux parties du jour alors que la hauteur de la couche-limite nocturne est constante et minimale.

Les concentrations aux limites (latérales et au sommet) sont nulles et le temps de mise à l'équilibre est de 24 heures (durée typique pour une simulation en Île-de-France).

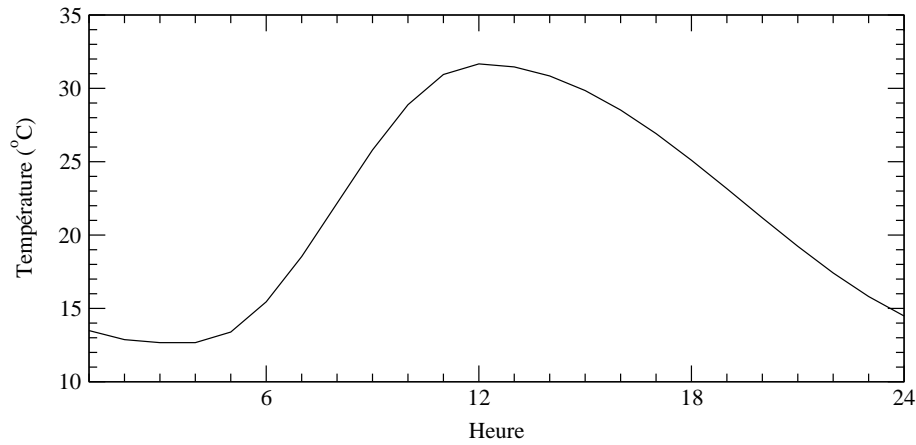


FIG. 3.6 – Évolution de la température à 2 mètres sur 24 heures.

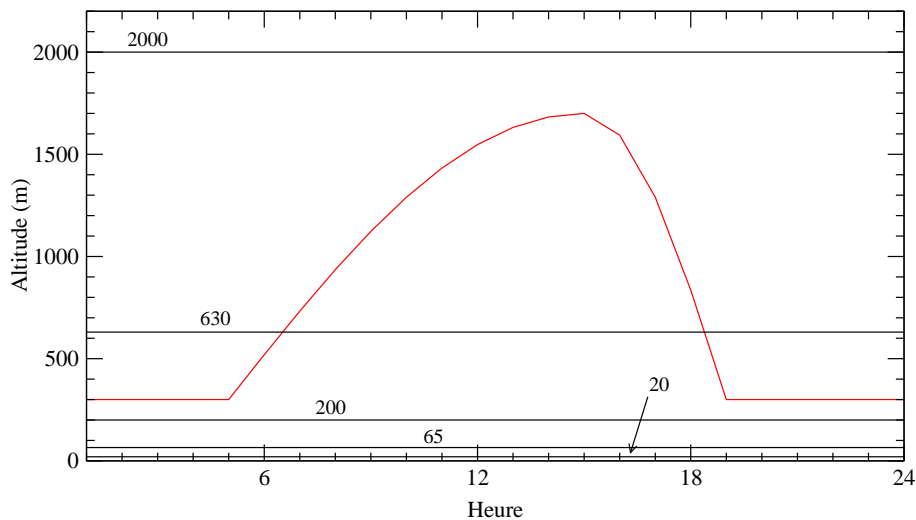


FIG. 3.7 – Niveaux verticaux et évolution de la hauteur de couche-limite en 24 heures.

### 3.6.1.2 Les émissions

Les émissions biogéniques sont quasiment nulles afin de mettre en évidence les relations entre émissions anthropiques et concentrations pour le monoxyde d'azote (NO est en effet une espèce émise anthropique et biogénique). Elles ne sont pas exactement égales à zéro pour éviter tout problème de calcul dans certaines routines traitant des émissions.

Le cadastre d'émissions comporte deux sources anthropiques, indiquées par les numéros 1 et 2 sur la Figure 3.8. Elles émettent les 16 espèces-modèle anthropiques. Les proportions des flux des différentes espèces sont réalistes. L'intensité maximale des flux d'émission des principales espèces sont de  $10^{14} \text{moléc.cm}^{-2}.\text{s}^{-1}$  pour CO,  $10^{13} \text{moléc.cm}^{-2}.\text{s}^{-1}$  pour  $\text{NC}_4\text{H}_{10}$  et NO,  $10^{12} \text{moléc.cm}^{-2}.\text{s}^{-1}$  pour  $\text{C}_2\text{H}_4$ ,  $\text{C}_2\text{H}_6$ ,  $\text{C}_3\text{H}_6$ , HCHO,  $\text{NO}_2$ , OXYL,  $\text{SO}_2$ . Pour chaque source, les émissions sont constantes et égales à  $E_{\text{max}}$  entre 7 et 11 heures et égales à  $E_{\text{max}}/10$  le reste du temps.

Le cadastre *a priori* est créé en sous-estimant les émissions «réelles» de  $\text{NO}_x$  de 20% pen-

dant les cinq heures d'intensité maximale. La proportion de 20% est considérée comme une estimation réaliste de l'incertitude du cadastre pour les émissions dues au trafic. La durée choisie correspond au nombre d'heures de pic de trafic pendant une matinée ou un après-midi. Deux inversions des émissions de  $\text{NO}_x$  perturbées sont effectuées, en conservant un rapport de 90% de  $\text{NO}$  pour 10% de  $\text{NO}_2$ . Dans les deux cas d'inversion, les émissions de  $\text{NO}_x$  sont inversées pendant une fenêtre temporelle de six heures, de 7 heures à midi. Le coefficient de correction qui devrait être obtenu est donc de 1,25 pour les cinq premières heures et 1 pour la dernière. Les différences entre cadastre *a priori* et cadastre «réel» amènent aux écarts entre concentrations de  $\text{NO}_x$  simulées représentées sur la Figure 3.8 pour  $\text{NO}$ .

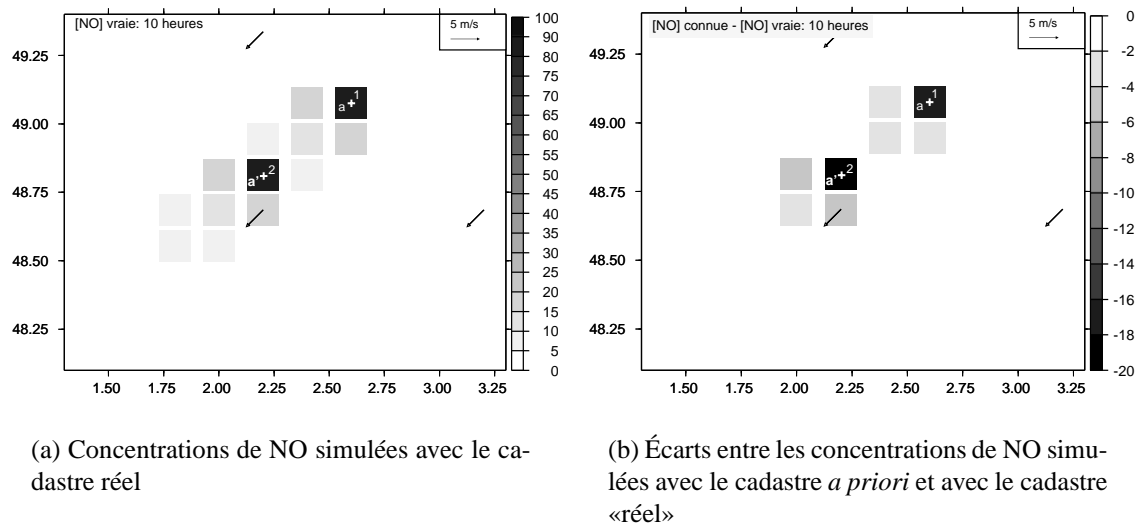


FIG. 3.8 – Concentrations surfaciques de  $\text{NO}$  et écarts (ppb) entre concentrations *a priori* et «réelles» à 10 heures.

## 3.6.2 Résultats

### 3.6.2.1 Cas formel 1 : précision théorique du système

Dans ce premier cas d'inversion, les deux stations de mesure (a) et (a') sont disponibles (indiquées sur la Figure 3.8). Ces stations fournissent des mesures «parfaites» de  $\text{NO}$ , l'inverse doit donc en principe permettre de retrouver les valeurs «réelles» des émissions. Il s'agit en effet d'un cas parfait pour l'inversion : toute l'information nécessaire est fournie sans erreur ni incertitude. Seules les erreurs dues au modèle direct peuvent dégrader les résultats.

L'optimisation converge et le coût s'annule puisque les contraintes sont parfaites. La Figure 3.9 résume les résultats pour les deux sources. Les coefficients de correction obtenus pour les deux sources sont de 1,25 pour les cinq heures perturbées et de 0,99 pour la dernière heure de la fenêtre. Les émissions «réelles» sont donc retrouvées.

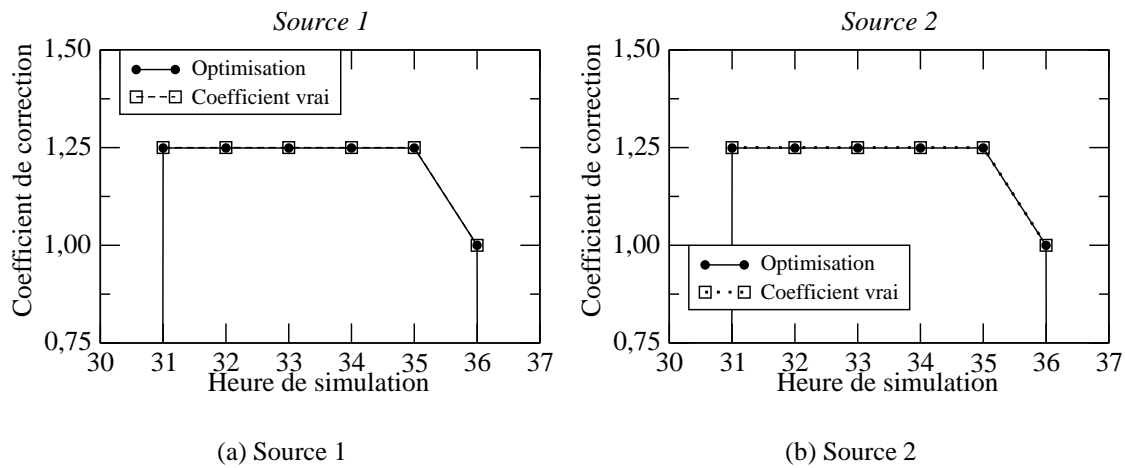


FIG. 3.9 – Coefficients correctifs pour les sources 1 et 2 à chaque heure de la fenêtre temporelle pour le cas formel 1.

### 3.6.2.2 Cas formel 2 : un exemple simple de manque d'information

Pour ce deuxième test, seule la station (a') est disponible et fournit des mesures parfaites. La seule information disponible intègre donc les effets des deux sources.

L'optimisation converge et le coût s'annule puisque les contraintes sont parfaites. La Figure 3.10 résume les résultats pour les deux sources.

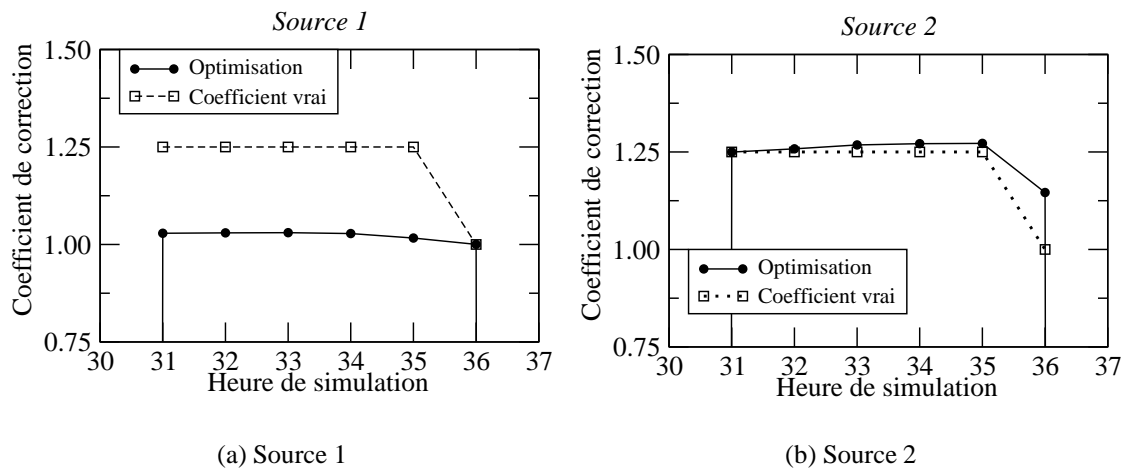


FIG. 3.10 – Coefficients correctifs pour les sources 1 et 2 à chaque heure de la fenêtre temporelle pour le cas formel 2.

La source 2, dans laquelle est placée la station (a') est presque parfaitement corrigée par l'inversion : le coefficient de correction pour les heures perturbées varie entre 1,25 et 1,27 et vaut 1,14 pour la dernière heure. L'intensité de la source 1 n'est quasiment pas modifiée par l'inversion : la correction pour les heures perturbées varie entre 1,01 et 1,03 et vaut 1 pour la dernière heure de la fenêtre temporelle.

Bien que la station (a') soit située en aval de la source 1, l'influence de celle-ci sur les



concentrations mesurées est faible à cause de la distance séparant la source et la station (voir la Figure 3.8) et parce que la source 2 a un impact très fort sur cette mesure. L'information fournie par la station (a') n'est donc pas pertinente pour la source 1 qui n'est donc presque pas modifiée par l'optimisation. Une légère surestimation de l'intensité de la source 2 est nécessaire pour simuler des concentrations exactement égales aux mesures à la station (a') puisqu'il faut compenser la sous-estimation de la source 1. À la dernière heure, la source 1 n'est plus sous-estimée mais la source 2 reste surestimée parce que la réponse des concentrations au changement d'émissions est décalée à cause du temps de transport.

### 3.7 Conclusion

L'inverse du modèle CHIMERE a été développé pour ce travail à partir du modèle direct version V200211K, de son adjoint et de l'optimiseur N1QN3. Cet outil doit permettre d'optimiser les paramètres d'entrée du modèle direct, notamment les émissions.

Une procédure permettant de générer des cas académiques a été mise en place afin d'évaluer les capacités de l'inverse. Les cas élaborés peuvent être très simples ou basés sur des cas réels.

Pour valider la gestion numérique de l'inverse, des cas académiques formels, très simplifiés afin d'assurer la rapidité des calculs ainsi que la mise en évidence des principaux aspects de l'inversion, ont été utilisés. Pour estimer la précision théorique du système, un premier cas, dans lequel toute l'information nécessaire était fournie, a été inversé. Les résultats parfaits à la précision de la machine près ont montré que la précision de l'inverse pouvait être très élevée.

Pour mieux comprendre les limites de l'outil d'inversion, l'inversion d'un deuxième cas, dans lequel l'information sur deux sources est intégrée en une seule mesure, a été effectuée. Cette première quantification des limitations de la méthode a mis en évidence deux effets complémentaires : (i) tout d'abord, la sur-correction de la source ayant l'impact le plus fort sur la mesure utilisée au détriment de la correction de la deuxième source, dont l'influence est nettement plus faible ; (ii) ensuite, la persistance de cette sur-correction dans le temps à cause du temps de transport entre la source et le point de mesure.

Finalement, l'utilisation de l'outil d'inversion pour l'étude de cas réels nécessite la mise en place d'une méthodologie permettant d'utiliser la meilleure information possible.

# Chapitre 4

## Méthodologie d'optimisation avec krigeage

**Utiliser les informations dans l'espace des concentrations** Le faible nombre de mesures disponibles par rapport à la dimension du problème et le manque d'information dans l'espace des paramètres (concernant le cadastre d'émissions) conduisent à rechercher un moyen d'obtenir les informations contenues dans l'espace des concentrations. Ceci semble possible grâce au **krigeage**, méthode d'hybridation entre modèle et mesures qui fournit un champ de concentrations **analysé** prenant en compte les informations fournies et par les mesures et par le champ simulé *a priori*. Si ce dernier présente déjà une structure validée, comme c'est le cas pour CHIMERE (voir notamment Schmidt et al. (2001) pour l'échelle continentale et [www.prevoir.org/fr/](http://www.prevoir.org/fr/) pour des statistiques quotidiennes), l'utilisation du champ analysé pour l'inversion permet de disposer de contraintes représentant la meilleure information possible dans l'espace des concentrations. De telles contraintes sont notamment adaptées à la représentativité du maillage.

### 4.1 Le krigeage

#### 4.1.1 Rappels théoriques

*Ces rappels utilisent les notations de Blond (2002a) et plus de détails sont disponibles dans (Blond 2002a; Blond et Vautard 2004; Blond et al. 2003) ainsi qu'en Annexe E.*

Le but du krigeage est de calculer les analyses  $y_{t,h}^a(\mathbf{s})$  d'un champ de concentrations  $y_{t,h}(\mathbf{s})$  pour tout point  $\mathbf{s}$ , le jour  $t$  ( $t = 1, \dots, T$ ) à l'heure  $h$  ( $h = 1 \dots, 24$ ). Un ensemble de  $p$  mesures de concentrations  $y_{t,h}^o(\mathbf{s}_k)$  avec  $k = 1, \dots, p$  et  $\mathbf{s}_k$  l'emplacement de la  $k^{\text{ième}}$  station de mesure et un champ de concentrations *a priori* (appelé aussi «first-guess»)  $y_{t,h}^b(\mathbf{s})$  sont disponibles. Les concentrations analysées  $y_{t,h}^a(\mathbf{s})$  sont alors données par :

$$y_{t,h}^a(\mathbf{s}) = y_{t,h}^b(\mathbf{s}) + \sum_{k=1}^p w_{t,h}^k(\mathbf{s})(y_{t,h}^o(\mathbf{s}_k) - y_{t,h}^b(\mathbf{s}_k)) \quad (4.1)$$

L'équation 4.1 montre donc que le champ analysé consiste en une correction du champ *a priori* par une combinaison linéaire des *innovations*,  $y_{t,h}^o(\mathbf{s}_k) - y_{t,h}^b(\mathbf{s}_k)$ . Les coefficients  $w_{t,h}^k(\mathbf{s})$  sont des poids à déterminer.

Dans le cadre du krigeage intrinsèque des innovations, les poids  $w_{t,h}^k(\mathbf{s})$  sont solutions du système Blond (2002a) :

$$\begin{cases} \forall k = 1, \dots, p, \\ \sum_{l=1}^p w_{t,h}^l(\mathbf{s}) [\gamma_{t,h}(\mathbf{s}_k, \mathbf{s}_l) + \sigma_{t,h}^{o2}(\mathbf{s}_k) \delta_{kl}] = \gamma_{t,h}(\mathbf{s}_k, \mathbf{s}), \\ \sum_{l=1}^p w_{t,h}^l(\mathbf{s}) = 1 \end{cases} \quad (4.2)$$

où  $\gamma_{t,h}(\mathbf{s}_k, \mathbf{s}_l)$  est la semi-variance des différences  $X(\mathbf{s}_k) - X(\mathbf{s}_l) = (y_{t,h}(\mathbf{s}_k) - y_{t,h}^b(\mathbf{s}_k)) - (y_{t,h}(\mathbf{s}_l) - y_{t,h}^b(\mathbf{s}_l))$  soit  $\gamma_{t,h}(\mathbf{s}_k, \mathbf{s}_l) = \frac{1}{2}E(X(\mathbf{s}_k) - X(\mathbf{s}_l))^2$  ( $E(\cdot)$  représente la moyenne) ;  $\sigma_{t,h}^{o2}(\mathbf{s}_k)$  est la variance de l'erreur d'observation. Cette dernière est supposée homogène et constante dans le temps, soit  $\sigma_{t,h}^{o2}(\mathbf{s}_k) = \sigma_h^{o2}$ . Pour calculer les semi-variances  $\gamma_{t,h}(\mathbf{s}_k, \mathbf{s}_l)$ , on utilise les semi-variances des différences spatiales des observations  $(y_{t,h}^o(\mathbf{s}_k) - y_{t,h}^b(\mathbf{s}_k)) - (y_{t,h}^o(\mathbf{s}_l) - y_{t,h}^b(\mathbf{s}_l))$  soit  $\gamma_{t,h}(\|\mathbf{s}_k - \mathbf{s}_l\|) = \gamma_{t,h}(r)$  avec  $r = \|\mathbf{s}_k - \mathbf{s}_l\|$ . La semi-variance  $\gamma_{t,h}(r)$  est obtenue grâce à un variogramme expérimental  $\gamma_{t,h}^{exp}$ . Celui-ci est calculé à partir de toutes les mesures disponibles à l'heure  $h$ , le jour  $t$ . Tout d'abord, les classes de distance sont construites avec un intervalle  $L$  et une tolérance  $\tau$ . Pour chaque classe, la variance  $\frac{1}{2}E(X^o(\mathbf{s}_k) - X^o(\mathbf{s}_l))^2$  est ensuite calculée. Enfin, le graphe des variances en fonction de la distance moyenne dans la classe est tracé et ajusté au semi-variogramme expérimental par un modèle exponentiel avec un «effet pépité» (noté  $NUG_h = \sigma_h^{o2}$ ).

#### 4.1.2 Mise en œuvre

La technique et le code de krigeage utilisés pour ce travail ont été élaborés par Nadège BLOND au cours de sa thèse, sous le nom de «méthode INK» (Blond 2002a; Blond et Vautard 2004; Blond et al. 2003). C'est un outil implanté dans la version de CHIMERE utilisée quotidiennement en Île-de-France par AIRPARIF pour réaliser les cartes de prévision (disponibles sur [www.airparif.org/page.php?article=regionale&rubrique=modelisation](http://www.airparif.org/page.php?article=regionale&rubrique=modelisation)).

La Figure 4.1 illustre un exemple d'analyse en Île-de-France, avec le réseau AIRPARIF, le modèle CHIMERE et la méthode de krigeage INK.

En Île-de-France, avec le réseau AIRPARIF, on utilise un intervalle  $L = 5$  km et une tolérance  $\tau = 4$  km pour les classes de distance. De plus, nous avons choisi pour  $\sigma_h^{o2}$  la variance maximale de toutes les mesures disponibles à l'heure  $h$ .

### 4.2 Krigeage et inversion itérative

L'utilisation des analyses fournies par le krigeage comme contraintes pour l'inversion permet d'obtenir d'une part les émissions optimisées et d'autre part les concentrations optimisées correspondant. On peut donc envisager d'utiliser ce nouveau champ de concentrations simulées pour effectuer une deuxième analyse puis une deuxième inversion des émissions. Des itérations successives de cycles krigeage-optimisation peuvent ainsi permettre d'affiner les résultats. Cependant, la question de la convergence et de l'optimalité d'une telle méthode itérative reste posée. L'Annexe F indique la piste suivie dans un cas linéaire simple. Dans le cadre de ce travail, la validation de la méthode a été abordée par le biais de cas académiques.

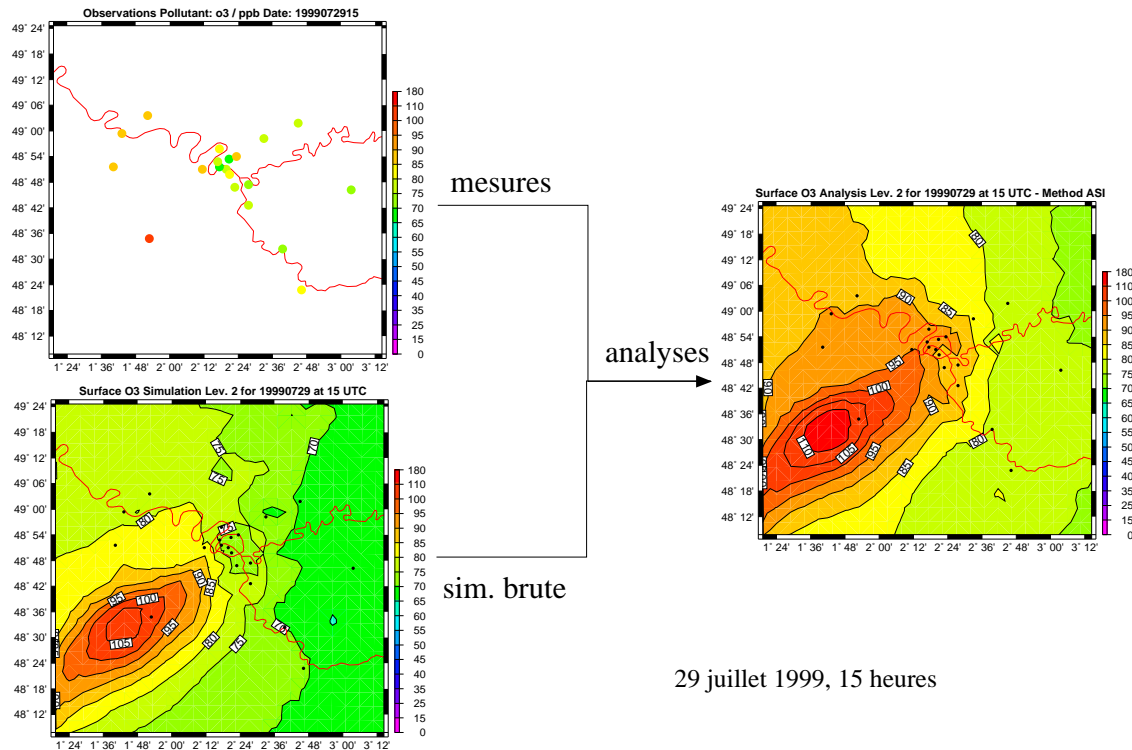


FIG. 4.1 – Exemple de résultats obtenus en Île-de-France avec le réseau AIRPARIF, le modèle direct CHIMERE et la méthode de krigeage INK.

### 4.3 Stratégie de validation : les cas académiques

#### 4.3.1 Principe

L'ajout du krigeage à l'élaboration des cas académiques décrite page 51 se fait comme indiqué sur la Figure 4.2. Le krigeage est effectué à partir des mesures et les concentrations *a priori*, obtenues avec le cadastre connu. Les concentrations analysées ainsi que les variances correspondant sont utilisées comme contraintes pour l'inversion.

#### 4.3.2 Cas académiques élaborés

Le cas académique-réel est élaboré à partir du 7 août 1998, qui fait partie de la POI2 d'ESQUIF (Menut 2003) (voir Annexe B). Lors de cette journée a eu lieu un épisode de pollution local, dû à un vent faible (moins de  $3 \text{ m.s}^{-1}$ ) avec une HCL faible (moins de 800 mètres), dans lequel les émissions locales jouent un rôle prépondérant. L'élaboration de cas académiques à partir de ce jour facilite donc la mise en évidence de l'influence des émissions et des différences entre cadastres.

Afin d'évaluer les possibilités et les limites de la méthodologie utilisant le krigeage pour l'inversion des émissions, il faut éviter autant que possible l'interaction avec d'autres limitations de la méthode.

Pour cela, on suppose tout d'abord que les sources d'émissions sont bien placées, seule leur intensité est perturbée et doit être inversée. De plus, aucun paramètre autre que les

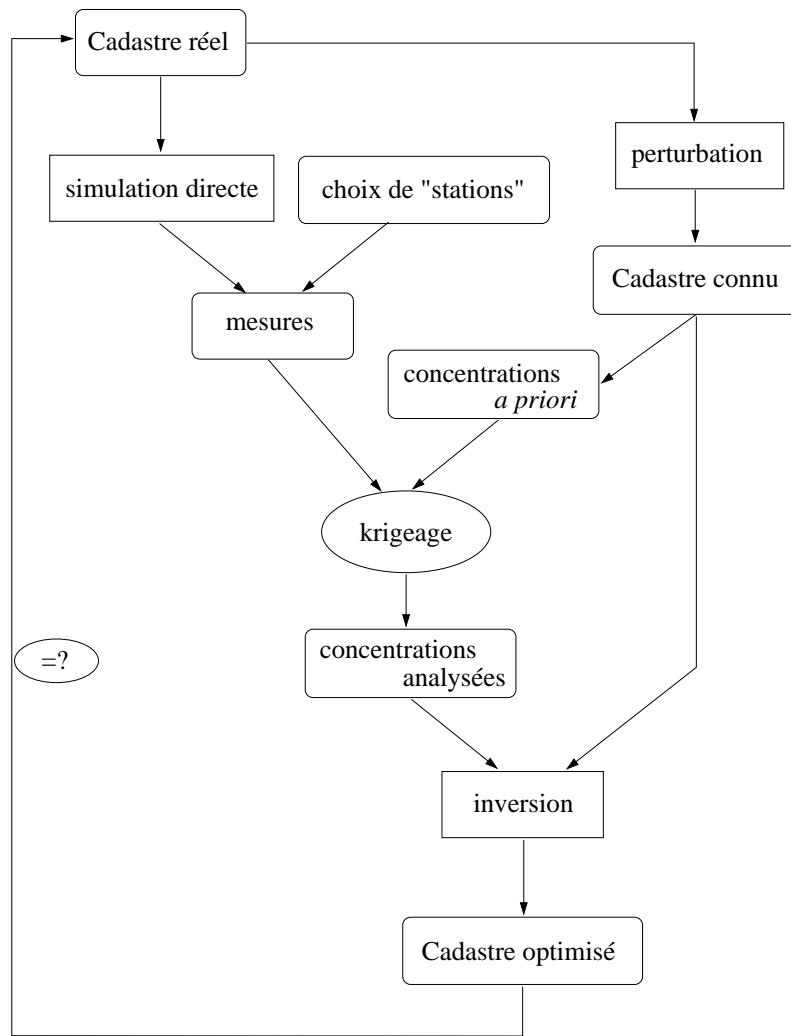


FIG. 4.2 – Principe d'élaboration d'un cas académique avec krigeage.

émissions n'est perturbé pour la simulation *a priori*. Une mauvaise estimation des paramètres météorologiques ou des variables diagnostiquées (comme la hauteur de couche-limite par exemple) ne peut donc pas interférer avec l'influence de l'erreur sur les émissions.

Le cadastre *a priori* est obtenu en perturbant, dans le cadastre «réel», les émissions de  $\text{NO}_x$  du pic de trafic matinal de 8 heures à 9 heures. Les flux concernés (les quatre plus intenses) sont indiqués sur la Figure 4.3. Le pourcentage de perturbation varie selon les cas :

- les cas «uN» correspondent à une sous-estimation des émissions de N% : pour u70, les émissions *a priori* ne représentent que 70% des émissions «réelles» ; sont aussi générés les cas u50 et u20
- les cas «oN» correspondent à une surestimation des émissions de N% : pour o20, les émissions *a priori* représentent 120% des émissions «réelles» ; sont aussi générés les cas o50, o70 et o100.

Les stations de mesure disponibles sont aux mêmes emplacements que celles d'AIR-

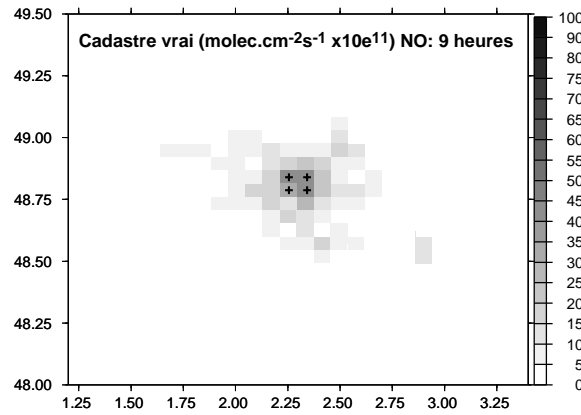


FIG. 4.3 – Cadastre d'émission «réel» : flux de NO à 9 heures le 7 août 1998.  $x$  = flux perturbé pour créer le cadastre *a priori*.

PARIF (cartes page 86). L'utilisation de réseau réel est simulée par l'interpolation des concentrations «réelles» aux emplacements exacts des stations de mesure. Les mesures générées ont donc la bonne représentativité par rapport au maillage utilisé. Les mesures de concentrations de NO ou d'ozone sont soit parfaites soit perturbées aléatoirement dans un intervalle de  $\pm 10$  ou  $\pm 15\%$  autour de la valeur «réelle». Les variances des mesures sont ensuite calculées grâce aux fonctions d'incertitude fournies par AIRPARIF.

### 4.3.3 Inversions

L'inversion porte sur les émissions de  $\text{NO}_x$  les plus intenses de 8 heures à 9 heures avec comme contraintes les mesures de concentrations de NO ou d'ozone de 9 heures à 17 heures.

Différents cas d'inversion sont comparés afin d'estimer l'influence de :

- la qualité des contraintes et notamment la différence entre l'utilisation des mesures et du krigeage
- la qualité du cadastre *a priori*
- la pertinence de l'information fournie par les contraintes (NO ou ozone).

## 4.4 Résultats

### 4.4.1 Définition du score

Un score est défini pour quantifier la qualité des résultats obtenus :

$$\text{score}_x = \left(1 - \frac{x_{\text{opt}} - x_{\text{réel}}}{x_{\text{apriori}} - x_{\text{réel}}}\right) \times 100$$

avec  $x_{\text{opt}}$  la valeur optimisée de  $x$ ,  $x_{\text{réel}}$  la valeur «réelle» et  $x_{\text{apriori}}$ , la valeur *a priori*,  $x$  étant un flux d'émission. La valeur du score correspond donc au changement en pourcentage de la différence entre le cadastre *a priori* et la «réalité» après l'optimisation. Un score positif indique que la différence avec la «réalité» a été réduite alors qu'un score négatif

indique qu'elle a été augmentée par l'optimisation. Une optimisation parfaite correspondrait donc à un score de 100% indiquant que la différence entre *a priori* et «réalité» a été annulée après optimisation.

Pour quantifier les résultats obtenus par l'optimisation des huit flux de la fenêtre temporelle, le score moyen et éventuellement la variance correspondant peuvent être calculés. La liste des cas inversés et leurs scores moyens est indiquée dans le Tableau 4.1 qui est discuté en détails dans les Sections suivantes.

Espèce utilisée	Type de contraintes	Qualité mesures $\epsilon_m$	qualité du cadastre <i>a priori</i> $\epsilon_E$						
			-70%	-50%	-20%	+20%	+50%	+70%	+100%
NO	mesures	0	53.8	35.4	x	x	x	x	x
		10	49.3	29.	x	-77.4	x	x	x
		15	57.	x	x	x	x	x	x
	krigeage	0	29.8	29.1	30.3	28.8	x	x	21.9
		10	31.6	31.4	32.7	26.8	25.7	25.3	21.
		15	29.7	30.2	42.8	17.4	26.2	22.2	20.9
ozone	mesures	0	x	x	x	x	x	x	x
		10	x	x	x	x	x	x	x
		15	x	x	x	x	x	x	x
	krigeage	0	x	8.3	7.3	5.7	0.6	x	x
		10	x	10.9	13.8	x	-4.1	x	x
		15	3.5	3.1	x	10.9	12.4	8.5	6.8

TAB. 4.1 – Scores moyens obtenus pour l'optimisation des émissions de  $\text{NO}_x$  avec des concentrations d'ozone ou de NO comme contraintes. x = convergence non atteinte.

#### 4.4.2 Nécessité du krigeage

L'utilisation de concentrations de NO comme contraintes pour l'inversion d'émissions de  $\text{NO}_x$  permet d'avoir un maximum d'informations. L'impact sur les concentrations de NO d'une erreur sur les émissions de  $\text{NO}_x$  est maximal à l'endroit et au moment de l'émission ainsi qu'immédiatement en aval, comme indiqué sur la Figure 4.7(b). Or, les stations de mesure se situent juste au-dessus et aux environs des émissions à inverser (Figure 4.7(b)). Cependant, le nombre total de contraintes pertinentes n'est pas suffisant : pour le cadastre o100, celui pour lequel l'écart avec la réalité est le plus grand, à 9 heures, moins de dix mesures sont situées en des endroits où cet écart est significatif (Figure 4.7(b)). Finalement, seuls six cas sur les vingt-et-un utilisant les mesures seules convergent contre dix-neuf sur les vingt-et-un utilisant le krigeage (Tableau 4.1, moitié supérieure). Les six cas utilisant les mesures seules qui atteignent la convergence correspondent à des écarts de -50 ou -70% entre le cadastre connu et le cadastre réel, les scores des cas u70 étant plus élevés. Un examen des résultats indique que l'optimum atteint est à chaque fois le même. Le score est donc d'autant plus grand que le point de départ était éloigné de la réalité. Lorsque des concentrations d'ozone sont utilisées comme contraintes, l'information portant sur les émissions de  $\text{NO}_x$  est plus dispersée dans le temps et l'espace. Une erreur sur les émissions de  $\text{NO}_x$  pendant le pic de trafic matinal amène une mauvaise estimation des concentrations d'ozone en plusieurs endroits : au-dessus des flux en question au moment

des émissions mais aussi dans le panache huit heures plus tard, comme indiqué sur la Figure 4.7(a). L'information concernant les  $\text{NO}_x$  émis le matin en ville est donc contenue principalement dans les concentrations d'ozone du panache de l'après-midi. Pour récupérer au moins une partie de cette information, le krigeage est nécessaire puisque très peu de mesures sont situées au bon endroit (Figure 4.7(a)). Finalement, avec des concentrations d'ozone comme contraintes, aucun cas n'utilisant que les mesures ne peut converger alors que treize des vingt-et-un cas utilisant le krigeage y parviennent (Tableau 4.1, moitié inférieure).

Les concentrations d'ozone n'étant pas linéairement liées aux émissions de  $\text{NO}_x$  (voir page 20), une grande perturbation des émissions de  $\text{NO}_x$  n'a pas nécessairement un fort impact sur les concentrations d'ozone. Comme indiqué sur la Figure 4.7(a), un doublement des émissions de  $\text{NO}_x$  induit des variations de moins de 15% dans les concentrations d'ozone de l'après-midi, à une heure où la sensibilité à ces émissions est pourtant élevée. D'un point de vue formel, l'utilisation du krigeage est donc nécessaire dans de nombreux cas pour que l'optimisation atteigne une solution mathématiquement acceptable.

#### 4.4.3 Impact du krigeage

La comparaison des cas dans lesquels les deux méthodes ont convergé permet de quantifier l'impact de l'utilisation du krigeage par rapport à celle des mesures seules. Les scores moyens obtenus dans ces cas (Tableau 4.1) sont reproduits dans le Tableau 4.2 avec les variances correspondant.

$\epsilon_E$	$\epsilon_m$	Type de contraintes	
		mesures	krigeage
-70%	0%	53.8 (10.9)	29.8 (7.7)
	10%	49.3 (9.9)	31.6 (2.7)
	15%	57. (14.4)	29.7 (8.7)
-50%	0%	35.4 (21.3)	29.1 (1.2)
	10%	29. (19.4)	31.4 (0.5)
+20%	10%	-77.4 (121.4)	26.8 (0.7)

TAB. 4.2 – Scores moyens (avec leur variance) obtenus pour l'optimisation des émissions de  $\text{NO}_x$  avec des mesures ou des analyses de  $\text{NO}$  comme contraintes.

L'écart moyen avec les émissions «réelles» est réduit de 45,1 à 59,3% en utilisant les mesures (mais il est augmenté dans un cas) et de 26,8 à 31,6% en utilisant le krigeage. Cependant, les variances plus élevées obtenues pour les cas utilisant les mesures seules (Tableau 4.2) indiquent que pour ceux-ci, les corrections ne sont pas réparties également sur tous les flux.

Le cas u50 avec des mesures de  $\text{NO}$  à  $\pm 10\%$  illustre la répartition des corrections obtenues par trois méthodes :

- avec les mesures seules
- avec les mesures et une matrice « de background » égale à l'identité par manque d'information
- avec le krigeage.



La matrice « de background » ajoute au coût, pour chaque flux à inverser, un terme de la forme  $b_{ii} \times (x_i(t) - x_{i,f}(t))^2$  (voir page 31) avec  $x_{i,f}(t) = 1 \forall i, t$  (il s'agit du coefficient multiplicatif *a priori*) et  $b_{ii} = 1 \forall i$  puisqu'on a choisi la matrice identité. Chaque composante du gradient comporte donc un terme supplémentaire de la forme  $2 \times b_{ii} \times (x_i(t) - 1)$ . Rappelons que la matrice « de background » a pour rôle de pénaliser les solutions qui s'éloignent trop des valeurs *a priori*.

La Figure 4.4 indique l'écart entre le flux optimisé et le flux réel (en %) pour les huit flux inversés (quatre flux à 8 heures et 9 heures). L'utilisation des mesures seules aboutit bien à une distribution des corrections très hétérogène en espace et en temps : dans la partie sud-ouest par exemple, le flux optimisé est sous-estimé de seulement 13,9% à 8 heures mais il est surestimé de plus de 37% à 9 heures. L'ajout de la matrice « de background » permet d'homogénéiser la répartition des corrections. Cependant, les meilleurs résultats sont obtenus avec le krigeage : les corrections sont réparties de façon homogène et l'écart moyen sur les huit flux inversés, qui est de 50% (en valeur absolue) pour le cadastre *a priori*, est réduit à 34,3%.

	mesures	mesures + B=I	krigeage
8 heures	-75.7   -56.3	-69.3   -51.4	-34.8   -35.5
	-13.9   -26.6	-16.6   -27.7	-26.5   -33.3
9 heures	-33.4   -33.1	-35.2   -38.3	-35.4   -34.0
	37.2   7.4	12.4   -14.6	-39.9   -35.1

FIG. 4.4 – Écart entre le flux optimisé et le flux réel (en % par rapport au flux réel) : comparaison des résultats obtenus en utilisant trois types de contraintes différents. Écart entre les flux *a priori* et les flux réels = -50%.

Une correction beaucoup plus grande de certains flux peut indiquer soit que le cadastre *a priori* n'est pas d'une qualité homogène soit que l'information n'est pas bien répartie. Les cas académiques permettent de comprendre ce qui se passe dans la seconde possibilité. Les flux eux-mêmes ne sont pas homogènes, ni dans l'espace ni dans le temps. Cependant, le but de l'inversion est d'optimiser le cadastre comme un tout. Les corrections doivent donc être réparties de façon aussi homogène que possible afin que la structure générale du cadastre, les rapports entre les différents flux (correspondant à des mailles urbaines ou rurales par exemple) soient respectés.

#### 4.4.4 Lien entre contraintes et espèce à inverser

##### 4.4.4.1 Concentrations de NO

La Figure 4.5 indique les scores obtenus pour l'ensemble des cas élaborés pour trois cycles krigeage - optimisation successifs (Tableau 4.1 pour le premier cycle). L'absence de score pour un des cycles signifie que la convergence n'a pas été atteinte dans le nombre d'itérations ou de simulations imparti (qui correspond à un temps de calcul de plusieurs heures). Cependant, lorsqu'un cycle n'a pas permis d'atteindre la convergence, il est tout de même possible de repartir du point atteint pour effectuer le cycle suivant.

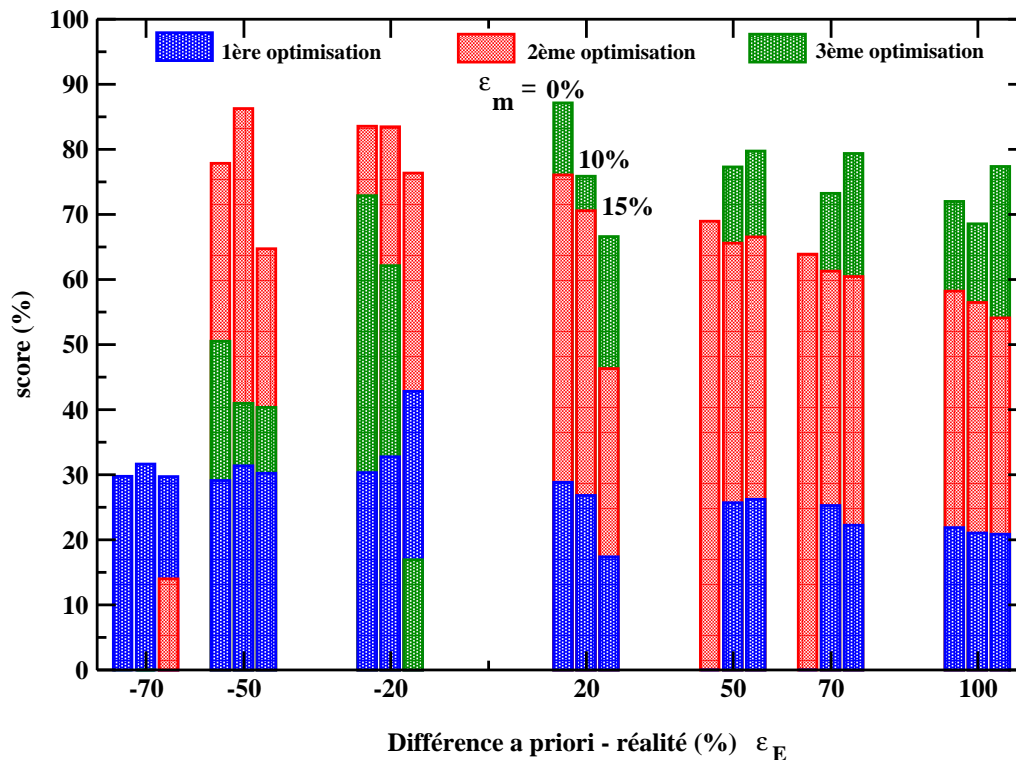


FIG. 4.5 – Scores moyens (%) après un à trois cycles krigeage-optimisation.

L'influence de la qualité des mesures n'est pas significative : quel que soit le nombre de cycles krigeage-optimisation considéré, une meilleure qualité des mesures ( $\epsilon_m$  plus petit) n'implique pas systématiquement un meilleur score. Il faut de plus noter que le principe du krigeage implique que toute mesure supplémentaire augmente la qualité des analyses. Si l'on considère les résultats de la première optimisation, l'influence de la qualité du cadastre *a priori* n'est pas limitante, du moins dans la fourchette de variation testée ici (de -70 à +100%), qui est raisonnable pour les émissions de  $\text{NO}_x$  en Île-de-France.

Le fait que le krigeage utilise les concentrations obtenues par la simulation directe «brute», effectuée avec le cadastre *a priori*, semble limiter les possibilités de l'optimisation. Après une optimisation, l'écart moyen avec la «réalité» est réduit de 17,4 à 42,8% (Tableau 4.1). Dix-neuf cas convergent après le deuxième cycle krigeage-optimisation. La qualité des résultats est améliorée pour 16 d'entre eux, pour lesquels l'écart avec la réalité est diminué de 46 à 86%, et la convergence est atteinte avec des scores supérieurs à 60% pour 2

autres cas. Un troisième cycle permet d'augmenter le score de 10 cas sur 15.

L'influence des différents cycles sur la qualité des contraintes obtenues est représentée sur la Figure 4.6 pour la maille correspondant au Sud-Ouest de la zone formée par les quatre flux. La comparaison des analyses montre qu'elles se rapprochent de la réalité à chaque cycle dans tous les cas ; le coût est bien diminué à chaque cycle, ainsi que la norme du gradient. Dans un cas réel, pour lequel on ne dispose pas de la comparaison à la réalité, il paraît donc difficile de prévoir si une optimisation supplémentaire conduira à un meilleur cadastre optimisé.

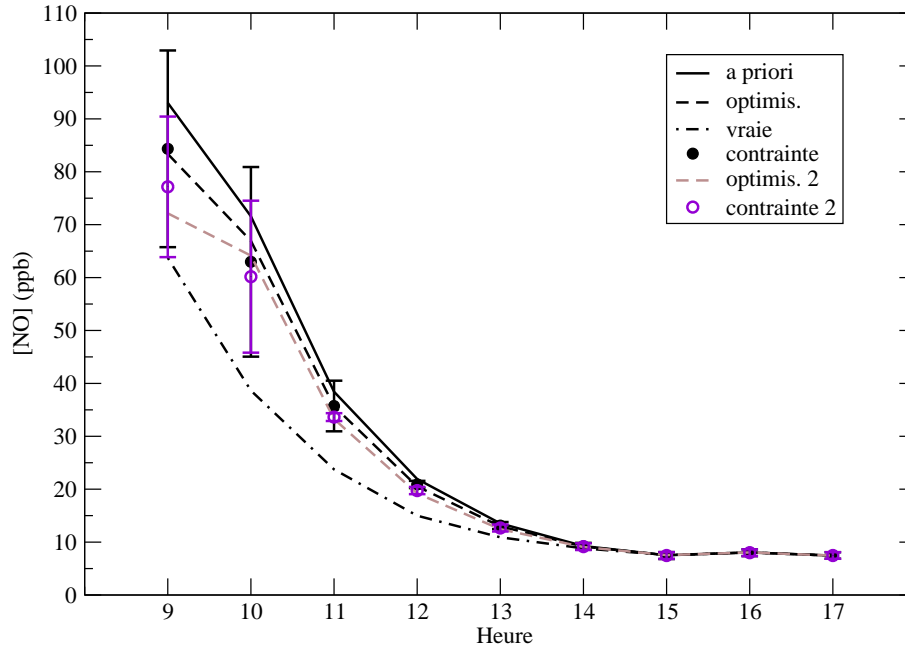


FIG. 4.6 – Évolution des concentrations de NO «réelles», *a priori* et optimisées pour  $\sigma 100$  avec les contraintes obtenues à partir de mesures à  $\pm 10\%$ , dans une maille située au-dessus d'un des flux inversés.

#### 4.4.4.2 Concentrations d'ozone

Les scores obtenus pour l'inversion des émissions de  $\text{NO}_x$  avec des concentrations d'ozone comme contraintes restent inférieurs à 15%, quel que soit l'écart entre les cadastres *a priori* et «réel» (Tableau 4.1).

Ces mauvais résultats s'expliquent par l'absence de mesures d'ozone à l'endroit et à l'heure auxquels la sensibilité aux émissions de  $\text{NO}_x$  est maximale. Comme indiqué sur la Figure 4.7(a), la direction du vent fait qu'il n'y a pas de mesure dans le panache. Cette situation est à comparer à celle des cas précédents : à l'endroit et à l'heure auxquels la sensibilité des concentrations de NO aux émissions de  $\text{NO}_x$  est maximale, c'est-à-dire au-dessus des émissions à 9 heures, de nombreuses mesures sont disponibles (Figure 4.7(b)). Les mesures d'ozone disponibles ne fournissent donc pas d'information pertinente, ou du moins pas assez, sur l'espèce à inverser c'est-à-dire les  $\text{NO}_x$  émis le matin en ville.

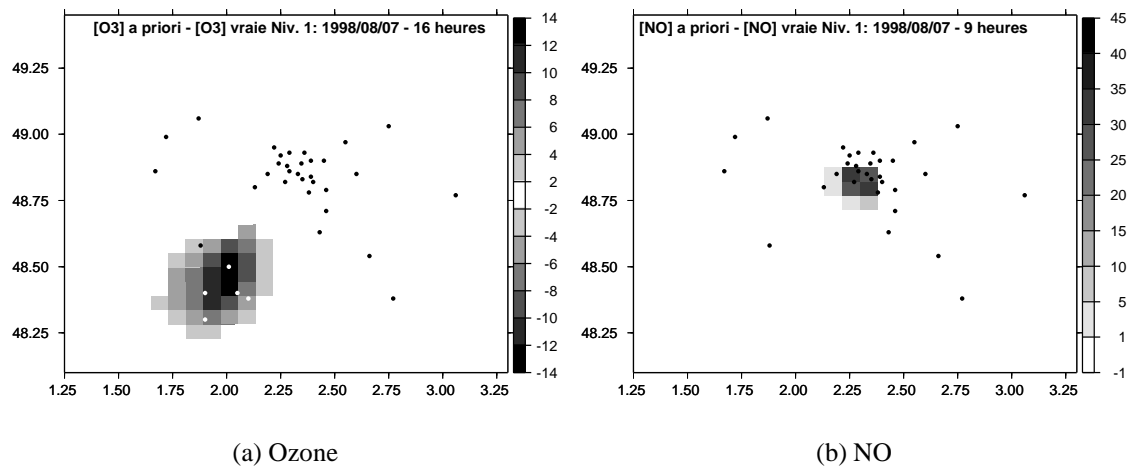


FIG. 4.7 – Écart (ppb) entre les concentrations simulées avec le cadastre *a priori* o100 et les concentrations «réelles» à l'heure du pic.

Ce manque d'information conduit le krigeage à peu modifier les concentrations *a priori* dans le panache : les analyses en sont donc très proches (voir Figure 4.8, à comparer à la Figure 4.6). Avec des contraintes peu différentes des valeurs *a priori*, l'inversion conduit à des flux optimisés proches des flux *a priori*.

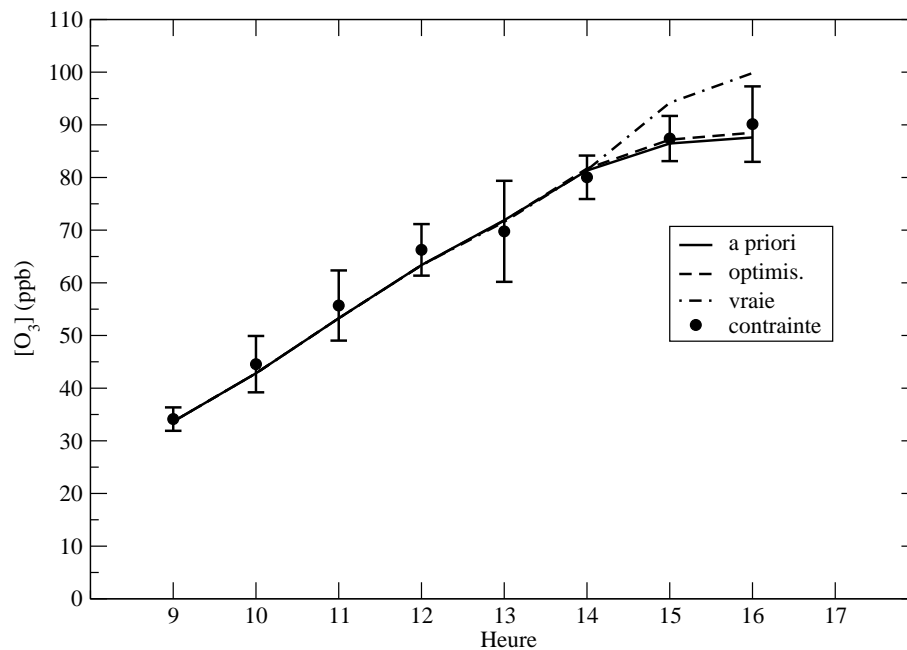


FIG. 4.8 – Évolution des concentrations d'ozone «réelles», *a priori* et optimisées pour o100 avec les contraintes obtenues à partir de mesures à  $\pm 15\%$ , au cœur du panache.

Des stations de mesures «virtuelles», indiquées sur la Figure 4.7 par des points blancs, ont été ajoutées dans le panache. Des mesures parfaites (à 0%) ont été générées avec ce nouvel ensemble de stations. Le premier cycle de krigeage et d'optimisation des sept cadastres

perturbés a été recalculé. Les résultats obtenus pour les quatre cas qui convergent après la première optimisation montrent une amélioration des scores pour  $u50$  (11.6 au lieu de 8.3) et  $u20$  (10.9 au lieu de 7.3) et une dégradation pour  $o20$  (4.1 au lieu de 5.7) et  $o50$  (-2.3 au lieu de 0.6). L'ajout d'informations dans le panache n'est donc pas aussi profitable selon l'erreur commise dans le cadastre connu. Dans notre cas, les stations supplémentaires permettent une meilleure correction d'une sous-estimation des émissions de  $NO_x$  le matin en ville alors qu'une surestimation des mêmes émissions reste mal corrigée.

**Cas des COV** L'inversion des émissions de COV nécessiterait l'utilisation de contraintes liées aux émissions de COV, par exemple des concentrations d'ozone. Comme indiqué dans le cas de émissions de  $NO_x$ , les informations contenues dans le panache d'ozone l'après-midi ne sont pas accessibles aux mesures. Or, dans le cas de l'Île-de-France par exemple, les flux les plus intenses de COV sont, comme pour les émissions de  $NO_x$ , localisés au centre du domaine (Figure 4.9) alors que leur influence sur les concentrations d'ozone est ressentie dans le panache de l'après-midi (Figure 4.10).

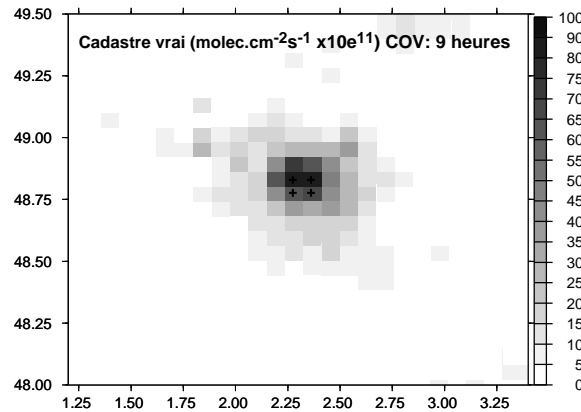


FIG. 4.9 – Cadastre d'émission «réel» : flux total de COV à 9 heures le 7 août 1998.  $x$  = flux perturbé pour créer le cadastre a priori.

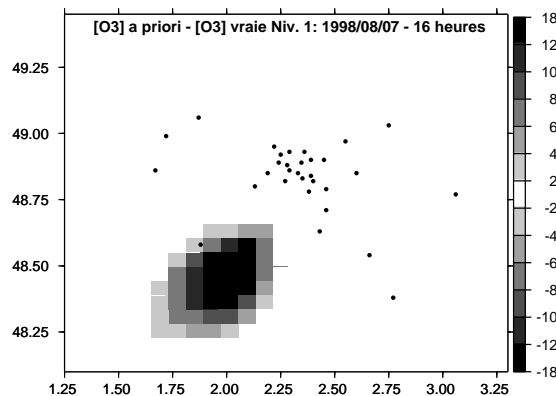


FIG. 4.10 – Écart (ppb) entre les concentrations d'ozone simulées avec le cadastre a priori  $o100$  pour les COV et les concentrations «réelles» à l'heure du pic.

## 4.5 Conclusion

L'utilisation des analyses comme contraintes pour l'inversion s'avère donc plus facile et sûre que celles des mesures seules ou d'une matrice «de background». La répartition spatiale et temporelle des corrections est plus homogène et cohérente pour l'ensemble du cadastre.

Lorsque l'information contenue dans les mesures est pertinente, il est possible d'affiner les résultats de l'inversion en effectuant des cycles krigeage-optimisation. Dans le cas où une optimisation n'a pas convergé, un cycle supplémentaire peut même permettre à l'inversion d'aboutir (cas o50 et o70 avec mesures parfaites). Les résultats peuvent cependant être dégradés dans certains cas.

L'utilisation du krigeage permet de répartir l'information disponible dans les mesures de façon à la rendre mieux exploitable par l'inverse mais, bien sûr, il n'y a pas création d'information. Il n'est donc pas possible de compenser un manque d'information dans les mesures. Dans le cas de base utilisé ici pour élaborer les cas académiques, aucune mesure n'est disponible dans le panache ; l'utilisation des concentrations d'ozone comme contraintes pour inverser les émissions de  $\text{NO}_x$  matinales n'est donc pas intéressante.

# Chapitre 5

## Agrégation spatiale et autres paramètres : compléter la méthodologie

### 5.1 L'agrégation spatiale

**Utiliser des informations *a priori* dans l'espace des paramètres** L'utilisation du krigage permet d'exploiter au mieux toutes les informations disponibles. La quantité totale d'informations reste cependant limitée et le problème de la grande dimension du problème par rapport à la quantité d'informations disponible demeure. Dans le cas de la pollution à l'échelle régionale, il est cependant particulièrement difficile d'ajouter de l'information dans l'espace des émissions (voir page 41).

Une technique simple d'**agrégation spatiale** dans l'espace des paramètres peut résoudre en partie le problème. En regroupant des paramètres pour lesquels on suppose que les corrections à apporter sont assez proches, il est possible de réduire la dimension du vecteur d'état (qui contient les coefficients de correction et non les paramètres eux-mêmes) c'est-à-dire celle du problème.

Tout ajout d'information dans l'espace des émissions, et l'agrégation spatiale en particulier, exige cependant d'effectuer des choix *a priori*, basés sur des hypothèses de travail. Aussi réalistes soient-elles au regard des données disponibles, leur validité est nécessairement discutable.

#### 5.1.1 Principe

Pour diminuer la dimension du problème, c'est-à-dire la taille du vecteur d'état, une solution simple consiste à regrouper les paramètres de telle sorte qu'une seule composante du vecteur d'état, c'est-à-dire un seul coefficient correctif, s'applique à plusieurs paramètres, par exemple plusieurs flux d'émission. Ces paramètres sont pour cela regroupés au sein d'une même **zone**.

On peut envisager de regrouper les paramètres spatialement et temporellement, des critères variés étant retenus pour définir l'emplacement et la taille des différentes zones d'intérêt (voir Heimann et Kaminski (1999) pour des exemples variés à l'échelle globale) : à l'échelle globale, Bousquet et al. (1999a) se basent sur la répartition géographique et les différents secteurs d'émissions, Prinn et Hartley (1995) recommandent d'utiliser les mesures de flux d'émission disponibles, enfin, Houweling et al. (1999) se basent sur la ré-

solution de leur modèle ; à l'échelle régionale, Mulholland et Seinfeld (1995) constituent des ensembles de sources particuliers. D'un point de vue théorique, le nombre de zones doit être égal au nombre de contraintes et en pratique est souvent inférieur (Heimann et Kaminski 1999). Le choix de zones très étendues, à l'échelle globale, peut introduire des erreurs importantes dues aux processus d'échelles inférieures à la taille des régions définies (Kaminski et al. 2001). Au sein de chacune des grandes zones, la distribution relative des flux d'émission est supposée connue *a priori* (Hartley et Prinn 1993). Seule l'intensité des émissions dans chaque zone est alors optimisée (Mahowald et al. 1997).

Cependant, dans notre cas, les mêmes raisons qui limitent les possibilités de mise en œuvre d'une matrice «de background» pertinente (voir page 41) obligent à procéder aussi simplement que possible, afin d'éviter d'avoir à faire des hypothèses trop rigides. C'est pourquoi il a été choisi de regrouper les paramètres spatialement mais non temporellement. On évite ainsi d'avoir à formuler des hypothèses sur le profil temporel des corrections. Les différentes zones sont alors définies comme des regroupements spatiaux de paramètres, par exemple en ensemble de flux d'émissions ruraux voisins.

L'**agrégation spatiale** ne doit pas effacer les disparités à l'intérieur de chaque zone afin de conserver la structure spatiale sous-jacente, notamment celle du cadastre dans le cas des flux d'émissions. On définit donc des zones, différentes pour chaque type de paramètre, à l'intérieur desquelles on applique le même coefficient à toutes les valeurs *a priori*. On fait donc simplement l'hypothèse que les paramètres regroupés doivent être corrigés de la même façon, dans la même proportion. Cette information peut être tirée des connaissances sur l'élaboration du cadastre. Par exemple, tous les flux d'émissions dus au trafic comportent probablement des erreurs du même type.

Cependant, la construction des zones doit également prendre en compte la sensibilité de la fonction coût aux paramètres regroupés. En effet, dans une zone  $z$ , pour les  $n$  valeurs d'un paramètre de type  $p$ , la composante en  $x_z$  du gradient est :

$$\begin{aligned}\frac{\partial J}{\partial x_z} &= \sum_{i=1}^n \frac{\partial J}{\partial p_i} \times \frac{\partial p_i}{\partial x_z} \\ &= \sum_{i=1}^n \frac{\partial J}{\partial p_i} \times p_i^0\end{aligned}$$

avec  $p_i^0$  la valeur *a priori* de  $p_i$ . Il faut donc regrouper les paramètres  $p_i$  de telle sorte que les termes en  $\frac{\partial J}{\partial p_i}$  ne se compensent pas. Sans cette précaution, il sera possible d'obtenir un gradient nul non représentatif du problème.

### 5.1.2 Mise en œuvre

L'agrégation spatiale a été conçue avant tout pour les flux d'émissions dans le cas d'un cadastre *a priori* sur lequel aucune autre information n'est disponible. Il a donc été supposé que des flux de même intensité étaient à corriger dans les mêmes proportions. Cette hypothèse est par exemple valable dans le cas où un même secteur d'activité est responsable de la plus grande partie des émissions d'une espèce, comme le trafic est responsable de presque toutes les émissions de  $\text{NO}_x$  en été en Île-de-France. Il est alors logique de supposer que les éventuelles erreurs dans la constitution du cadastre sont les mêmes pour tous les flux urbains ou ruraux par exemple.



Il n'est pas possible d'effectuer systématiquement des calculs de sensibilité pour délimiter des zones cohérentes pour chaque type de paramètre. En effet, cela nécessiterait un calcul adjoint complet par maille du domaine, ce qui annulerait le bénéfice en temps de calcul de la méthode de l'adjoint. Cependant, il est possible de déterminer grossièrement des zones suffisamment cohérentes d'après la physique du cas étudié, approche qui dépend donc principalement de la configuration du site d'étude.

Dans les cas où une zone de sources très intenses existe, on peut distinguer, en plus de cette zone, trois régions, schématisées sur la Figure 5.1 :

- une région en amont de la zone source, particulièrement sensible aux conditions aux limites et aux émissions locales
- une région située directement sous le vent de la source, dans le panache, qui est donc particulièrement sensible à l'intensité de la source
- une région en aval de la source mais qui n'est pas située directement dans son panache.

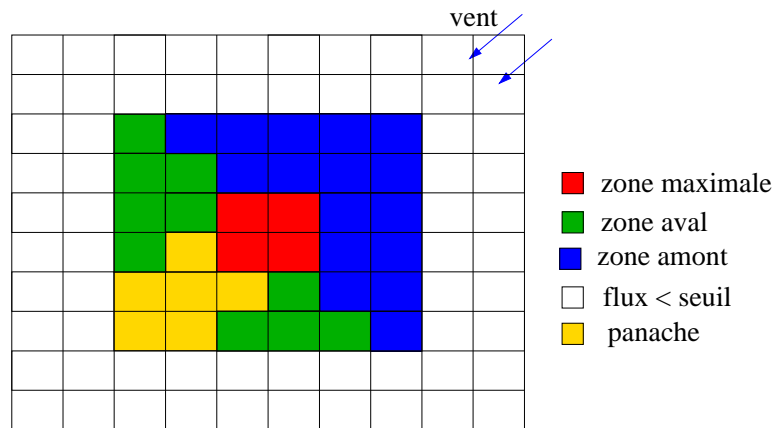


FIG. 5.1 – Schéma d'un découpage en grandes régions selon la direction du vent.

Finalement, pour les flux d'émissions, la construction des zones se fait en quatre étapes :

Étape 1 : un seuil est imposé afin de ne pas traiter les flux d'émission très faibles (Figure 5.1). Dans le cas des  $\text{NO}_x$  par exemple, on peut ne s'intéresser qu'aux flux du même ordre de grandeur que le flux maximum ou prendre en compte les flux de l'ordre de grandeur inférieur.

Étape 2 : la zone d'émission maximale est déterminée : c'est la zone qui comprend le flux maximal et ceux qui sont compris dans la tolérance  $\tau_E$  choisie. La tolérance  $\tau_E$  est le pourcentage maximum d'écart souhaité entre le flux maximal et les flux regroupés dans la même zone : celle-ci englobe donc le flux maximal  $E_{max}$  et tous ceux qui sont au moins égaux à  $(1 - \tau_E) \times E_{max}$ .

Étape 3 : les zones «amont», «aval» et «de panache» (de la zone d'émission maximale) sont déterminées à partir de la direction du vent dans la zone maximale. Cette étape de calcul simple est valable pour la région parisienne mais sera éventuellement à élaborer davantage dans le cas de topographies plus complexes.

Étape 4 : à l'intérieur de chaque zone de l'écoulement, les flux sont regroupés selon la tolérance  $\tau_E$  voulue. Dans le panache par exemple, peuvent se trouver des mailles

contenant des flux d'émissions assez intenses de banlieue et des mailles rurales dans lesquelles les flux sont faibles. Si  $\tau_E$  n'est pas trop grande, ces deux types de flux sont finalement regroupés dans des zones différentes.

Le nombre de zones obtenu dépend du pourcentage maximum d'écart souhaité. Plus la tolérance est grande, plus le nombre de zones est petit. Les quatre étapes de la construction des zones sont répétées à chaque heure à inverser, c'est-à-dire à une résolution temporelle correspondant aux variations d'intensité des flux d'émissions et à celles de la direction du vent. Il est important de noter que l'agrégation spatiale effectuée est donc dynamique : les zones définies varient selon les heures en fonction de la direction du vent et des variations d'intensité des flux d'émission.

Pour d'autres types de paramètres (météorologiques par exemple), la construction des zones devra être adaptée selon la structure du champ : par exemple, pour un champ de température assez homogène dans tout le domaine, une seule zone suffit (voir page 78).

### 5.1.3 Validation par des cas académiques

Il s'agit ici d'estimer dans quelle mesure la diminution de la dimension du problème permet au critère de convergence défini précédemment d'être atteint et si l'agrégation spatiale élaborée pour les flux d'émissions permet d'obtenir la précision voulue.

#### 5.1.3.1 Cas académiques élaborés

Le même principe que dans le cas du krigeage (voir page 60) est utilisé pour élaborer des cas académiques-réels sur la base du 7 août 1998, comme décrit page 60. Les flux perturbés sont toujours les émissions de  $\text{NO}_x$  à 8 heures et 9 heures mais leur nombre varie. Les différents cadastres *a priori* sont indiqués dans le Tableau 5.1.

Nom du cas	$\epsilon_E$ (%) $\frac{a\text{ priori}-\text{réel}}{\text{réel}}$	Flux perturbés	
		Intensité ( $\text{moléc.cm}^{-2}.\text{s}^{-1}$ )	Nombre
o50-58	+50	$> 1.10^{12}$	58
o20-58	+20	$> 1.10^{12}$	58
o50-127	+50	$> 5.10^{11}$	127
o20-127	+20	$> 5.10^{11}$	127
o-58	+50/+40	$> 1.10^{12}$	58

TAB. 5.1 – Flux perturbés et différence en % entre les cadastres *a priori* et «réel» entre 8 et 9 heures. Le nombre de flux perturbés est le total sur les deux heures.

#### 5.1.3.2 Inversions

L'inversion porte sur les émissions de  $\text{NO}_x$  de 8 heures à 9 heures avec comme contraintes les analyses de concentrations de NO de 9 heures à 17 heures, élaborées à partir de mesures «parfaites» ( $\epsilon_m = 0$ ). Différents cas d'inversion sont comparés afin :

- de montrer l'impact de la diminution de la taille du problème

- de déterminer s’il existe une dimension optimale par rapport à la quantité d’information disponible et au critère de convergence choisi
- d’estimer la précision atteinte par l’agrégation spatiale mise au point pour les flux d’émissions.

### 5.1.3.3 Résultats

#### Nécessité de la réduction de la dimension

Le cas o50-58 a été inversé en réduisant peu à peu la dimension du problème c’est-à-dire en procédant à des regroupements au sein de zones de plus en plus larges. Les résultats obtenus sont indiqués dans le Tableau 5.2. La dimension est de 57 et non 58 avec une tolérance  $\tau_E = 0\%$  car deux des flux les plus intenses sont égaux. Les zones obtenues sont représentées sur la Figure 5.2 pour une seule heure, la dimension étant le nombre total de zones à inverser pendant les deux heures.

Tolérance (%)	Dimension	Score (%)
0	57	x
10	33	45,66
20	21	54,64

TAB. 5.2 – Scores moyens obtenus pour l’optimisation des émissions de  $NO_x$  avec des concentrations de  $NO$  comme contraintes, selon la dimension du problème (cas o50-58).  $x$  = convergence non atteinte.

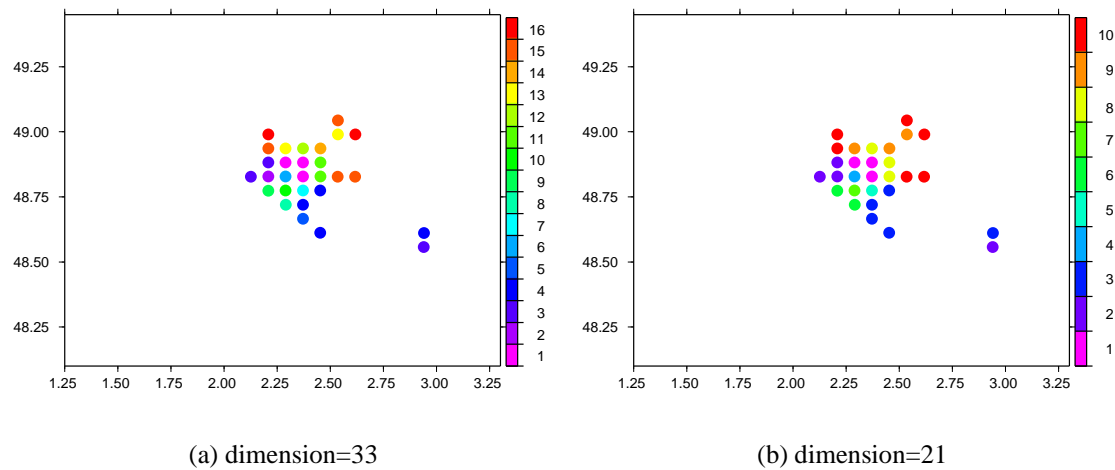


FIG. 5.2 – Zones obtenues pour le cas o50-58 à 8 heures (29 flux), selon la tolérance  $\tau_E$  choisie.

La réduction de la dimension du problème de 57 à 33 permet bien d’atteindre la convergence. Rappelons que le nombre de mesures disponibles est de 24 par heure pour une fenêtre temporelle de 8 heures. Les scores obtenus sont satisfaisants puisque la différence

entre les cadastres *a priori* et «réel» est réduite de plus de 45% après la première optimisation. Le principe théorique et le traitement numérique de l'agrégation spatiale sont donc validés.

Dans ce cas, plus la dimension est réduite, meilleur est le résultat de l'optimisation : le score augmente de 9 points quand la tolérance  $\tau_E$  augmente de 10 à 20%. Malheureusement, cet effet positif de la réduction de la dimension n'apparaît que parce que, dans ce cas particulier, tous les flux regroupés sont véritablement à corriger dans la même proportion. Dans un cas réel, l'hypothèse de l'uniformité de la correction n'est jamais exacte et constitue une limite de la méthode.

### Une dimension optimale ?

Plusieurs cas ont été comparés afin de déterminer s'il existe une dimension permettant d'obtenir systématiquement la convergence de l'optimisation, la quantité d'information étant toujours la même. Le Tableau 5.3 indique pour divers cas le nombre total de flux concernés et la dimension la plus grande permettant d'obtenir la convergence de l'optimisation.

Cas	Nombre de flux	Dimension
o50-58	58	33
o20-58	58	33
o50-127	127	14
o20-127	127	26

TAB. 5.3 – Dimension permettant d'obtenir la convergence pour l'optimisation des émissions de  $NO_x$  avec des concentrations de NO comme contraintes, selon le nombre de flux du problème.

Bien que les cadastres *a priori* ne comportent pas la même erreur, les cas o20-58 et o50-58 convergent tous les deux pour une dimension de 33. Il semble donc que, pour un problème donné (même «physique» et mêmes informations), la dimension permettant à l'optimisation de converger ne dépend que de la quantité d'informations disponible. Cependant, les cas o20-127 et o50-127 infirment cette possibilité. Pour le même nombre de flux à inverser, un cadastre *a priori* surestimé de seulement 20% permet de n'avoir à réduire la dimension que de 127 à 26 (soit environ 4,9 fois) pour atteindre la convergence contre une réduction de 127 à 14 (soit plus de 9 fois) pour un cadastre *a priori* surestimé de 50%. Il apparaît donc que la dimension maximale permettant à l'optimisation de converger ne dépend pas seulement de la quantité d'informations disponible mais aussi de la qualité du cadastre *a priori* et des contraintes. Cette conclusion attendue a des conséquences pratiques sur la procédure d'inversion. En effet, pour un problème donné, il sera nécessaire de procéder de façon assez empirique à des réductions successives de la dimension, jusqu'à atteindre celle qui permettra la convergence de l'optimisation.

### L'hypothèse d'homogénéité de la correction

Afin de déterminer l'influence de l'hypothèse de l'uniformité de la correction sur la structure des résultats, le cas o-58 a été créé en perturbant différemment les flux des mailles

paires et impaires (Figure 5.3 : mailles paires à +50%, mailles impaires à +40%). La dimension a dû être réduite à 9 pour que l'optimisation converge ; des flux perturbés différemment ont été regroupés.

La différence entre le cadastre optimisé et la réalité conserve la structure en damier imposée au cadastre *a priori* (Figure 5.3). L'homogénéité de la correction au sein d'une zone ne permet pas d'obtenir une erreur plus homogène pour le cadastre optimisé que pour le cadastre *a priori*.

Le score moyen global atteint cependant 56,18%, soit une différence moyenne avec la «réalité» réduite de plus de la moitié. Le score moyen des flux surestimés de 50% est de 54,53 et celui des flux surestimés de 40% est de 58,22. Ces résultats après un cycle krigeage-optimisation sont très satisfaisants. Si l'erreur du cadastre optimisé n'est pas homogène, la cohérence est maintenue puisque les écarts ont été réduits quasiment dans la même proportion pour les deux perturbations.

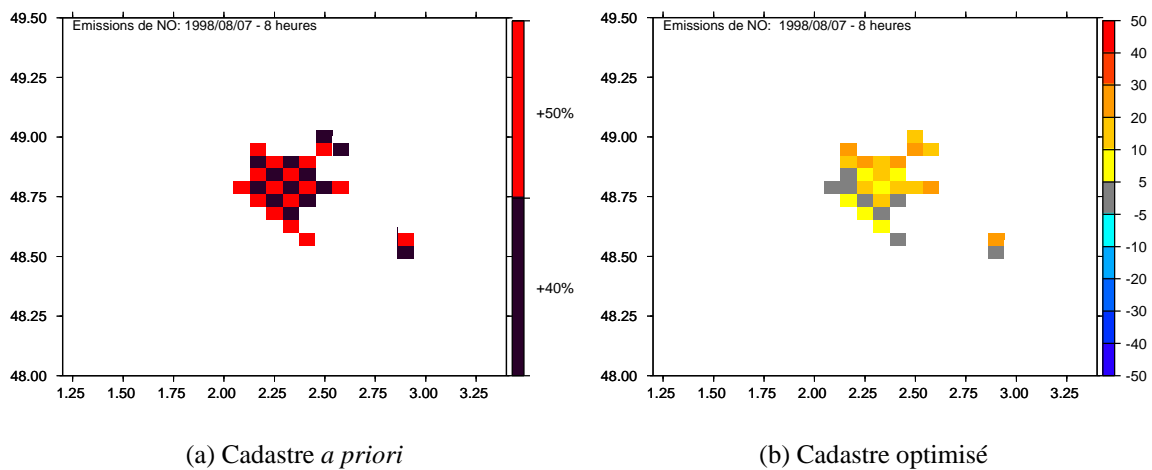


FIG. 5.3 – Différence (en %) avec le cadastre «réel» de NO pour le cas o-58 à 8 heures.

#### 5.1.3.4 Conclusion

L'agrégation spatiale utilisée ici est basée sur l'hypothèse que les paramètres regroupés sont à corriger dans la même proportion. Le principe général de l'agrégation spatiale peut être mis en œuvre de diverses façons. Le choix présenté ici est très simple et adapté à une région comme l'Île-de-France : les zones regroupent des flux d'émissions d'intensités proches selon la physique de la situation qui permet de prendre en compte le signe de la sensibilité du coût à ces émissions.

Des cas académiques ont permis de valider la gestion numérique et de montrer l'intérêt de la réduction de la dimension du problème. Les scores obtenus après un seul cycle krigeage-optimisation sont très satisfaisants, supérieurs à 45% dans les exemples détaillés. L'agrégation spatiale élaborée ici permet de diminuer la dimension du problème de telle sorte que le critère de convergence est atteint et que les flux d'émission sont inversés avec la précision voulue.

Cependant, il est également apparu que la dimension permettant à l'optimisation de converger ne pouvait être déduite uniquement de la quantité d'information disponible mais du

problème dans son ensemble (la qualité du cadastre *a priori* joue notamment un rôle important). En pratique, cela signifie que l'inversion d'un cas réel demandera plusieurs essais, avec des réductions successives de la dimension.

## 5.2 Prise en compte de paramètres autres que les émissions

### 5.2.1 Nécessité

Les émissions font partie des paramètres auxquels les concentrations simulées sont les plus sensibles. Mais selon le cas étudié, d'autres paramètres peuvent jouer un rôle aussi important (Menut 2003). Or, l'hypothèse inhérente à la méthodologie choisie attribue aux paramètres à inverser tout l'écart entre concentrations mesurées et simulées. L'inversion des émissions seules aboutit alors au report de tout l'écart entre mesures et modèle sur les flux optimisés : il y a sur-correction des émissions pour compenser les erreurs sur tous les autres paramètres. Le cadastre d'émissions obtenu n'est alors plus indépendant de l'outil de modélisation utilisé.

Bien que le but de ce travail soit l'optimisation des émissions, il pourrait donc être nécessaire de prendre en compte dans l'inversion au moins les paramètres autres que les émissions ayant une influence comparable sur les concentrations simulées. Il est en effet impossible d'inverser tous les paramètres d'entrée : le problème aurait une dimension énorme. Il faut donc effectuer un choix parmi les paramètres les plus influents selon le cas considéré.

Les paramètres sélectionnés peuvent ensuite être pris en compte de plusieurs façons. Ils peuvent être inversés en même temps que les émissions mais ce choix conduit à augmenter la dimension du problème et nécessite éventuellement l'ajout d'hypothèses, notamment par le biais d'une matrice «de background» (voir page 78). Dans le cas de paramètres comportant un biais connu ou une incertitude bien déterminée, il peut donc être plus intéressant de prendre en compte l'erreur qui leur est due avant l'inversion elle-même (voir page 81).

### 5.2.2 Mise en œuvre dans l'inverse : la matrice «de background»

L'utilisation d'une matrice dite «de background», qui était à éviter pour les émissions, est souhaitable dans le cas de variables météorologiques. En effet, d'une part, les valeurs *a priori* de ces variables ont une incertitude connue et, d'autre part, CHIMERE est un modèle «off-line». Cela signifie donc que les valeurs optimisées de variables météorologiques doivent rester assez proches des valeurs *a priori* pour assurer la cohérence des diagnostics effectués (voir page 41).

La température est un paramètre qui peut être inversé sans conduire à une trop grande augmentation de la dimension du problème. En effet, dans une région à la topographie assez simple (pas de côtes, pas de relief trop prononcé, une seule grosse agglomération), on peut envisager d'agréger en seulement deux zones tout le domaine, une zone constituée par l'îlot de chaleur urbain et la deuxième par le reste de la région.

Le vecteur d'état  $x$  comprend alors des coordonnées supplémentaires  $x_{temp,r}$  qui sont les coefficients à appliquer à toutes les températures de chaque zone  $r = 1, 2$ . La fonction coût est donc enrichie d'un terme de la forme :

$$J_B(x(t)) = \sum_{r=1}^2 w_{B,r} X_{temp,r}^2$$

avec  $w_{B,r}$  le poids pénalisant l'écart entre la valeur optimisée et la valeur *a priori*. Comme

$$x_{temp,r} = \frac{temp_i}{temp_{i,fg}} \quad \forall \text{ la maille } i \text{ dans la zone } r$$

avec  $temp_{i,fg}$  la valeur *a priori* de la température, on a :

$$\begin{aligned} \frac{\partial J}{\partial x_{temp,r}} &= \frac{\partial J_R}{\partial x_{temp,r}} + \frac{\partial J_B}{\partial x_{temp,r}} \\ &= \sum_{i=1}^N \frac{\partial J_R}{\partial temp_i} \times \frac{\partial temp_i}{\partial x_{temp,r}} + \frac{\partial w_{B,r} X_{temp,r}^2}{\partial x_{temp,r}} + \frac{\partial w_{B,s} X_{temp,s}^2}{\partial x_{temp,r}} \\ &= \sum_{i=1}^N \underbrace{\frac{\partial J_R}{\partial temp_i}}_{adjoint} \times temp_{i,fg} + \underbrace{2 \times X_{temp,r} \times w_{B,r}}_{dJ_{B,r}} \frac{\partial x_{temp,r}}{\partial x_{temp,r}} + dJ_{B,s} \underbrace{\frac{\partial x_{temp,s}}{\partial x_{temp,r}}}_{=0} \end{aligned}$$

Rappelons que les termes notés «adjoint» sont calculés par l'adjoint en même temps que les termes correspondant  $\frac{\partial J_R}{\partial flux_i}$ . D'où finalement :

$$\frac{\partial J}{\partial x_{temp,r}} = \sum_{i=1}^N temp_{i,fg} \times \frac{\partial J_R}{\partial temp_i} + dJ_{B,r}$$

La composante supplémentaire du gradient est donc de la même forme que les autres sauf qu'il faut ajouter le terme  $dJ_{B,r}$ . *Remarque* : il n'y a pas de changement dans les autres composantes du gradient puisque la température ne dépend pas des émissions à notre échelle.

**Exemple d'inversion de la température et des émissions** Le modèle simple de cas académiques décrit page 53 comporte un cadastre *a priori* dans lequel les émissions de  $NO_x$  des deux sources sont sous-estimées de 20% de 7 heures à 11 heures, tous les autres paramètres étant «parfaits». Afin d'estimer la précision théorique de l'inversion de la température en même temps que les émissions, un champ de température *a priori* est élaboré en sous-estimant la température de 10% (en Kelvin, donc de façon très exagérée) autour des sources pendant les cinq heures d'intensité maximale.

Les inversions des émissions de  $NO_x$  et des températures perturbées sont effectuées pendant une fenêtre temporelle de six heures, de 7 heures à midi. Le coefficient de correction qui devrait être obtenu est donc, pour les cinq premières heures, 1,25 pour les émissions et 1,11 pour les températures et 1 pour la dernière heure. Les différences entre cadastre *a priori* avec météorologie *a priori* et cadastre «réel» avec météorologie «réelle» amènent aux écarts entre concentrations de  $NO_x$  simulées représentées sur la Figure 5.4 pour  $NO$ .

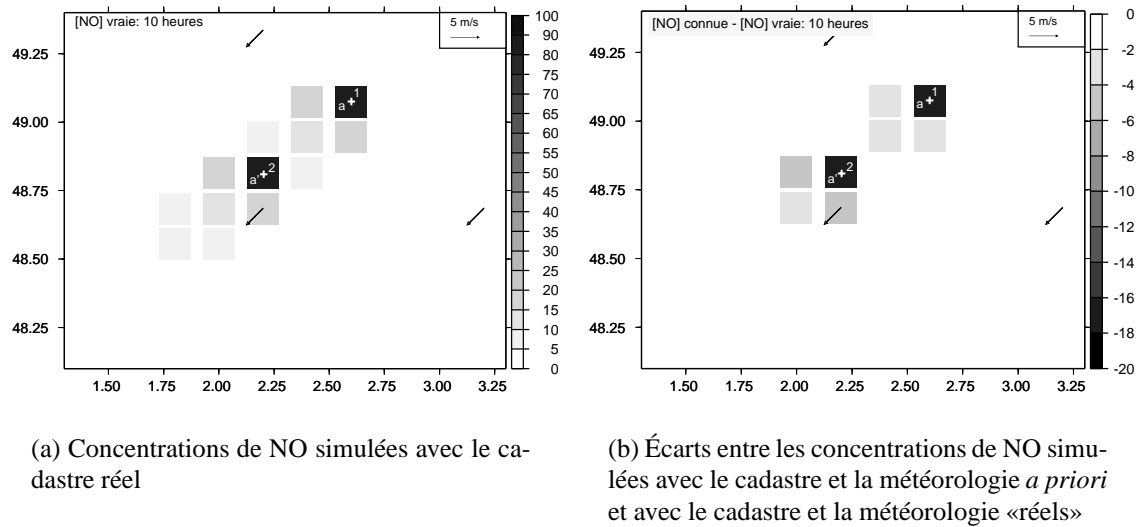


FIG. 5.4 – Concentrations surfaciques de NO et écarts (ppb) entre concentrations *a priori* et «réelles» à 10 heures.

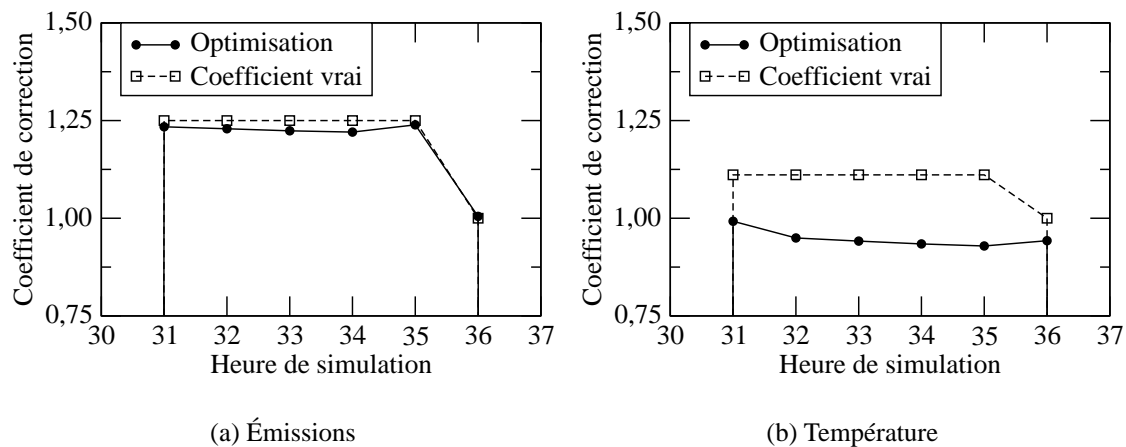


FIG. 5.5 – Coefficients correctifs pour les émissions (pour la source 2) et les températures à chaque heure de la fenêtre temporelle quand les flux d'émissions sont indépendants.

Les deux stations de mesure (a) et (a') sont disponibles (indiquées sur la Figure 5.4) et fournissent des mesures «parfaites» de NO. Dans les deux cas inversés, les deux températures à optimiser sont regroupées en une seule zone.

Dans le premier problème inversé, les deux flux d'émissions sont inversés indépendamment. Seul l'ajout de la matrice «de background» compense donc l'augmentation de la dimension provoquée par l'ajout de la zone des températures. La Figure 5.5 résume les résultats pour les émissions et les températures. Bien que les émissions optimisées des deux sources soient quasiment égales aux émissions «réelles», ces résultats ne sont pas aussi bons que ceux de l'inversion des sources seules décrite page 55. Les températures optimisées ne sont même pas corrigées dans le bon sens : les coefficients correctifs sont inférieurs à 1. Dans ce problème, l'information ajoutée grâce à la matrice «de background»



ne suffit donc pas à obtenir un cadastre et un champ de température optimisés corrects. Dans le second problème, la dimension est donc réduite en regroupant les deux flux d'émissions. La Figure 5.6 résume les résultats pour les émissions et les températures. Les émissions «réelles» sont parfaitement retrouvées après l'inversion. Les températures optimisées sont presque égales aux températures «réelles» sauf pour la deuxième heure (correspondant à 8 heures) où le problème de sous-estimation persiste. De plus, il est nécessaire, pour obtenir la convergence, d'effectuer davantage d'itérations que dans le cas de l'inversion des émissions seules.

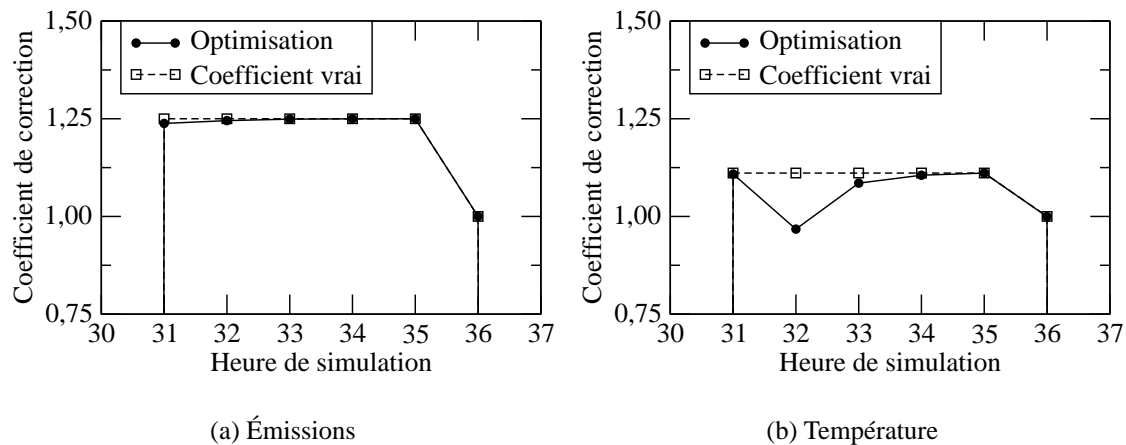


FIG. 5.6 – Coefficients correctifs pour les émissions et les températures à chaque heure de la fenêtre temporelle quand les flux d'émissions sont regroupés.

Finalement, l'inversion de la température en même temps que les émissions est possible mais ne semble pas, dans ce cas académique, très fiable.

## 5.2.3 Mise en œuvre en amont de l'inverse

### 5.2.3.1 Correction de biais

Les conditions aux limites et la direction du vent sont des paramètres susceptibles de comporter des biais non négligeables mais qui sont, dans certains cas, repérables en comparant les concentrations mesurées et simulées. Une station située au bord du domaine en amont des sources principales peut suffire à indiquer un biais dans les conditions aux limites. Les stations situées dans le panache d'une source très intense, par exemple Paris en Île-de-France, permettent de déterminer si la direction du vent est correcte. Quand le champ de vent est de direction et de module constants sur une vaste étendue, dans des situations de turbulence faible, il est possible de rectifier ces biais en les soustrayant directement des paramètres *a priori*.

**Exemple de panache urbain avec un biais dans la direction du vent** La Figure 6.9 page 99 indique les concentrations d'ozone simulées par CHIMERE pour le 7 août 1998 à 15 heures en surface ainsi que des mesures de stations du réseau AIRPARIF (voir le Chapitre 6 pour les détails des simulations et des mesures). L'intensité du panache est

effectivement reproduite mais il est légèrement décalé vers le Sud par rapport aux observations (ESQUIF 2001; Hodzic 2002; Menut et al. 2000), comme l'indique par exemple la mesure à 112 ppb qui est à la frange du panache modélisé au lieu d'être au centre. Ce glissement est dû au fait que les champs météorologiques d'ECMWF, qui ne sont pas finement résolus, sont linéairement interpolés.

### 5.2.3.2 Encadrement pour une incertitude connue

Dans le cas de la hauteur de couche-limite, il est plus difficile de détecter et corriger un biais. Si l'on ne souhaite pas inverser ce paramètre, il est tout de même possible d'obtenir un ordre de grandeur de son influence sur les résultats de l'inversion. En effectuant plusieurs inversions d'un même cas avec une HCL *a priori* légèrement modifiée, on obtient un jeu de cadastres optimisés que l'on peut comparer.

**Exemple d'inversion avec incertitude sur la HCL** Dans le cas des émissions de  $\text{NO}_x$  inversées avec des contraintes de NO, on peut s'attendre à ce que la relation entre la variation de la HCL et l'écart sur les corrections des émissions soit linéaire. En effet, la hauteur de couche-limite détermine le volume dans lequel les émissions sont diluées.

Le cas o50-58 a été simulé et inversé avec une HCL surestimées de 10% dans tout le domaine. Le score obtenu est de 40,6% au lieu de 45,7% lorsque la HCL est «réelle». Il faut remarquer que dans un cas réel, on ne peut dire quelle inversion est plus proche de la réalité. Le score plus faible obtenu avec la HCL perturbée est dû au fait que la surestimation du volume dans lequel elles sont diluées compense en partie la surestimation des émissions. Avec un plus grand volume de dilution des émissions, les concentrations de NO sont déjà plus proches de la «réalité» avec le cadastre *a priori*. La différence moyenne entre les corrections obtenues avec la HCL «réelle» et avec la HCL surestimée de 10% est de 10,2 points, ce qui correspond à la relation prévisible quasiment linéaire entre l'écart sur les corrections des émissions de NO et la variation de la HCL.

On dispose donc de deux méthodes pour prendre en compte les erreurs sur des paramètres météorologiques avant une inversion : si un biais a été détecté, on peut l'éliminer avant la simulation directe ; si leur incertitude est bien connue, on peut faire varier les valeurs *a priori* et comparer les cadastres optimisés.

## 5.3 Conclusion : la méthode complète

### 5.3.1 L'inversion

La Figure 5.7 résume la méthode d'inversion complètement élaborée. Il faut d'abord effectuer le krigeage qui fournit les contraintes pour l'inversion. Les stations du réseau disponibles pour l'espèce (ou les espèces) choisie(s) fournissent les mesures. Les paramètres connus (émissions, météorologie) sont utilisés pour effectuer une simulation directe qui donne les concentrations *a priori*. Le krigeage utilise les concentrations mesurées et simulées pour élaborer le champ de concentrations analysées (analyses et variances).

Il faut ensuite préparer le vecteur des paramètres que l'on veut inverser. Pour les émissions, il peut être nécessaire de procéder à une agrégation spatiale, qui tient compte des

caractéristiques du cadastre connu. Il faut effectuer un choix parmi les autres paramètres, comme la météo, afin de ne garder que les plus pertinents.

L'inverse utilise comme contraintes les analyses du krigeage et optimise les paramètres contenus dans le vecteur d'état. On obtient finalement un cadastre optimisé pour les émissions et éventuellement des corrections sur les autres paramètres choisis.

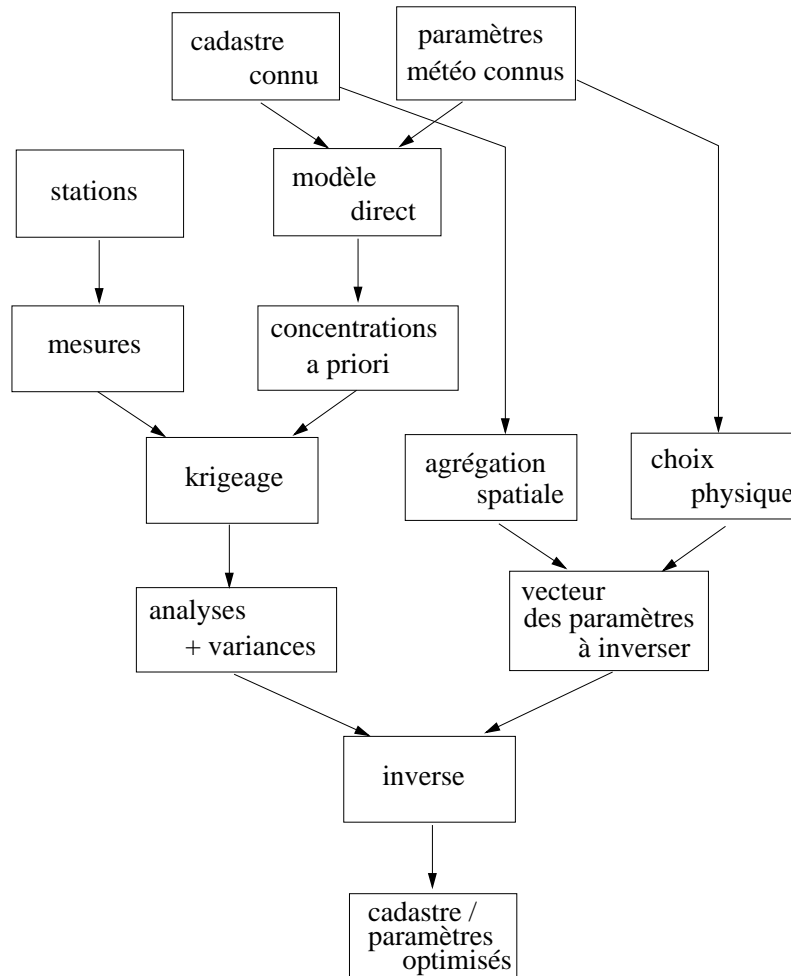


FIG. 5.7 – Tous les éléments de notre méthode d'inversion.

### 5.3.2 Stratégie générale

L'optimisation des émissions par modélisation inverse nécessite une stratégie générale autour de l'inversion proprement dite.

Il faut tout d'abord identifier les problèmes rencontrés dans la simulation des concentrations dans la région d'étude, par exemple, des biais importants ou une mauvaise reproduction des pics de  $\text{NO}_x$ .

Il faut ensuite déterminer les paramètres ayant la plus grande influence sur ces écarts. L'étude des particularités de la physique du site ou des études de sensibilités permettent en général de retenir quelques paramètres-clefs comme les émissions de certaines espèces, les conditions aux limites ou le champ de vent.

Pour choisir les paramètres-clefs à prendre en compte, il faut connaître au moins un ordre de grandeur des diverses incertitudes. Les paramètres ayant une incertitude faible ne doivent en effet pas être modifiés. Cette sélection doit permettre de retenir les paramètres qui ont à la fois une grande influence sur les concentrations et une grande incertitude.

Afin de choisir la meilleure façon de prendre en compte les paramètres autres que les émissions, il est important de connaître la manière dont s'exerce leur influence sur les concentrations. Les conditions aux limites, par exemple, ont sur la plupart des espèces un impact linéaire : les concentrations importées dans le domaine s'ajoutent à celles qui résultent de la production locale. Il est donc plus simple et facile de «débiaiser» les conditions aux limites en amont de l'inverse plutôt que de les inverser.

Une fois que les paramètres à inverser ont été choisis, on peut effectuer l'inversion elle-même, comme décrit page 82 et exploiter les résultats, c'est-à-dire le cadastre et éventuellement les autres paramètres optimisés.

# Chapitre 6

## Application à des cas réels

### 6.1 Le cadre de l'étude : l'Île-de-France

Il a été choisi de mener les premières études en Île-de-France, région qui présente plusieurs avantages de ce point de vue. Tout d'abord, sa topographie simple avec un relief faible, son éloignement des côtes contribuent à créer des champs dynamiques plus homogènes et donc plus facilement reproductibles lors des simulations que dans une région comportant des reliefs prononcés ou une côte favorable aux phénomènes de brise. De plus, le réseau de mesure AIRPARIF est très complet (extension, densité, espèces mesurées) et la campagne de mesure ESQUIF des étés 1998 et 1999 permet de disposer de huit cas de pollution, d'une durée de deux à cinq jours chacun (Annexe B), pour lesquels les impacts des différentes incertitudes ont déjà été étudiés (Beekmann et Derognat 2003; Menut et al. 2000; Menut 2003; Vautard et al. 2003).

### 6.2 Position du problème

Afin de déterminer les problèmes que pose le cadastre d'émissions en Île-de-France, deux études ont été effectuées. La première porte sur des processus d'émission qui sont absents du modèle : les émissions tridimensionnelles des avions dans la couche-limite. La seconde est une étude de comparaison entre les mesures et le modèle pour la campagne ESQUIF.

#### 6.2.1 Impact des émissions 3D

Les émissions utilisées dans les modèles se présentent sous forme de flux surfaciques, qui correspondent aux sources d'émissions les plus importantes, comme le trafic routier ou les industries. Or, d'autres sources existent, non plus surfaciques mais en trois dimensions : les avions de ligne, au décollage et à l'atterrissage, dans les trois aéroports civils de taille importante situés en Île-de-France (Le Bourget, Roissy-Charles-de-Gaulle et Orly). Même si ces avions ne sont évidemment pas une source primaire majeure en milieu urbain, il paraît important de quantifier l'impact de ces émissions tridimensionnelles sur les concentrations de polluants dans la couche-limite dans et autour de l'agglomération avant de procéder à l'inversion des émissions.

Ce travail préliminaire est décrit dans Pison et Menut (2004), reproduit en Annexe G. Les résultats obtenus montrent que les émissions dues au trafic aérien au sein de la couche limite ont un impact non négligeable sur les concentrations d’ozone dans l’agglomération (Figure 6.1). L’impact global net de ces émissions sur les concentrations d’ozone est négatif, ce qui indique que les aéroports accentuent la titration de l’ozone par NO. Cet impact est confiné au voisinage des aéroports (Figure 6.1), à cause de la rapidité des réactions chimiques combinée à des vitesses de vent faibles (situation la plus fréquente lors des cas de pollution observés en Île-de-France).

Finalement, il a été montré que la prise en compte de ces émissions permettait de mieux reproduire certains phénomènes observés pendant la campagne ESQUIF (dont des processus de titration de l’ozone par NO, mesurés par les avions lors de leur passage à la verticale du Bourget). Cependant, la procédure d’implémentation des émissions tridimensionnelles est particulièrement lourde à gérer dans le modèle de chimie-transport, pour un impact sur les résultats qui n’explique pas complètement les différences entre les mesures et le modèle : l’erreur majeure n’est pas due à ces émissions particulières. Dans un premier temps, il a donc été décidé de ne pas prendre en compte ces émissions aéroportées lors des procédures d’inversion. On inversera donc les émissions anthropiques **surfaciques**.

## 6.2.2 Comparaisons entre les mesures et le modèle

Cette étude doit permettre de détecter et d’estimer les erreurs (par exemple des biais systématiques) du modèle dues notamment à la météorologie et aux émissions.

### 6.2.2.1 Les données

#### Les mesures de routine AIRPARIF

En 1999, les mesures AIRPARIF sont effectuées dans 35 stations représentatives de trois types d’environnements dits urbain, rural ou «trafic» et 4 stations complémentaires de type «observation» ; en 1998, dix stations supplémentaires étaient disponibles. Les mesures de routine pour l’ozone, le monoxyde d’azote et le dioxyde d’azote sont effectuées avec une résolution temporelle d’une heure mais toutes les mesures ne sont pas disponibles dans toutes les stations. L’emplacement des stations, leur type et les mesures disponibles sont indiquées sur la Figure 6.2.

Les trois types de station (décrits page 23) n’ont pas la même représentativité. Les stations «trafic» sont représentatives d’un très petit volume, tout au bord de la voie sur laquelle elles sont placées ; c’est pourquoi elles font l’objet d’une étude spéciale dans ce qui suit. Les stations «observation» ne sont pas utilisées car leur représentativité n’est pas bien contrôlée.

Dans les stations équipées, le dioxyde d’azote n’est pas mesuré directement. La méthode employée conduit à surestimer la concentration de NO<sub>2</sub> car d’autres composés azotés tels que le PAN, le HONO, le HNO<sub>3</sub> et certains nitrates organiques y sont inclus. Les mesures de NO<sub>2</sub> fournies sont donc plus proches des concentrations de NO<sub>y</sub> simulées. Dans la suite, seules les observations d’ozone et de monoxyde d’azote sont prises en compte.

Les concentrations mesurées sont fournies en  $\mu\text{g.m}^{-3}$ . Comme les mesures sont effectuées près du sol, il est possible de les convertir facilement en ppb (l’unité de référence du modèle) en utilisant les coefficients constants de conversion, qui sont typiquement de  $\frac{1}{1,95}$

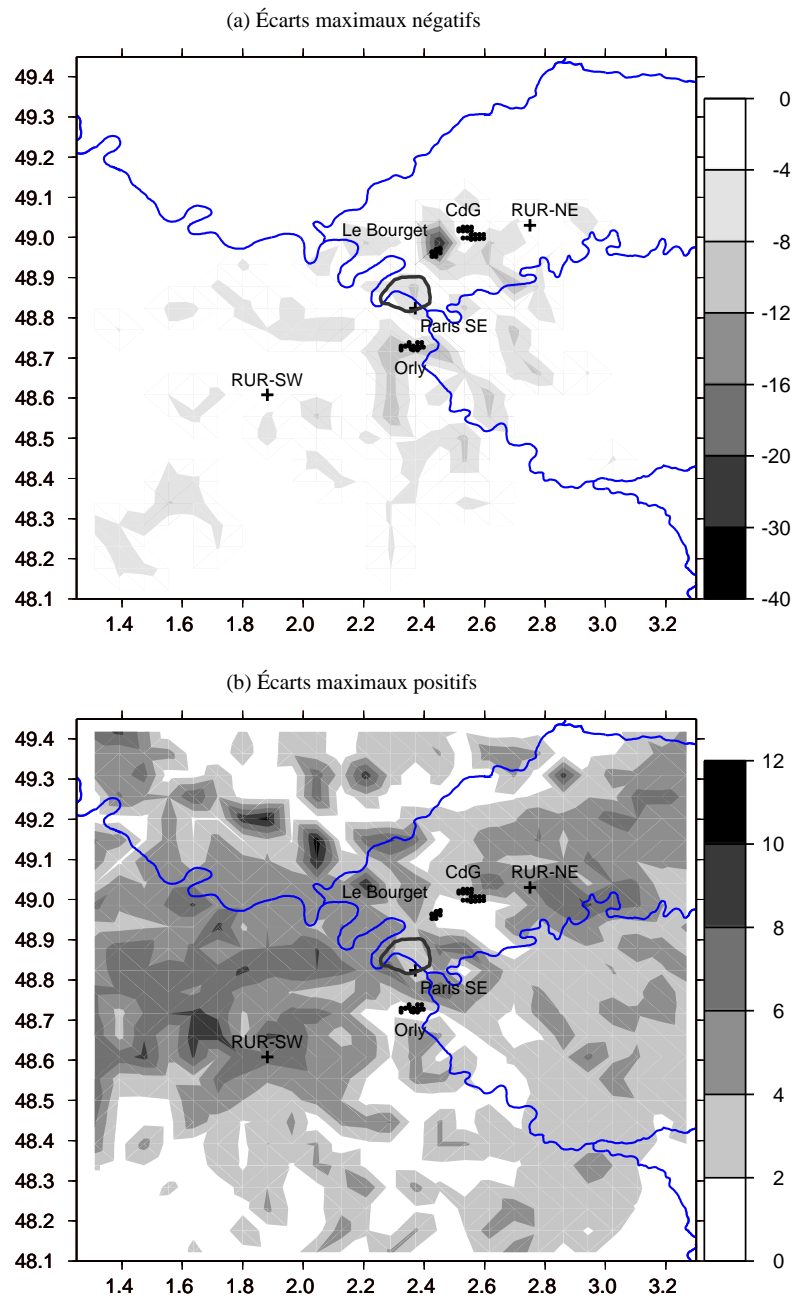


FIG. 6.1 – Cartes composites d'écarts (ppb) entre les concentrations d'ozone de la simulation prenant en compte les émissions des avions jusqu'à 3000 mètres et de la simulation de référence sans émission tridimensionnelle : pour chaque maille, l'écart maximal (négatif et positif) obtenu sur toute la durée de la simulation de la POI 2 à tous les niveaux est représenté. L'échelle est choisie de telle sorte qu'on ne prend pas en compte les 10% les plus faibles. CdG = aéroport Charles-de-Gaulle.

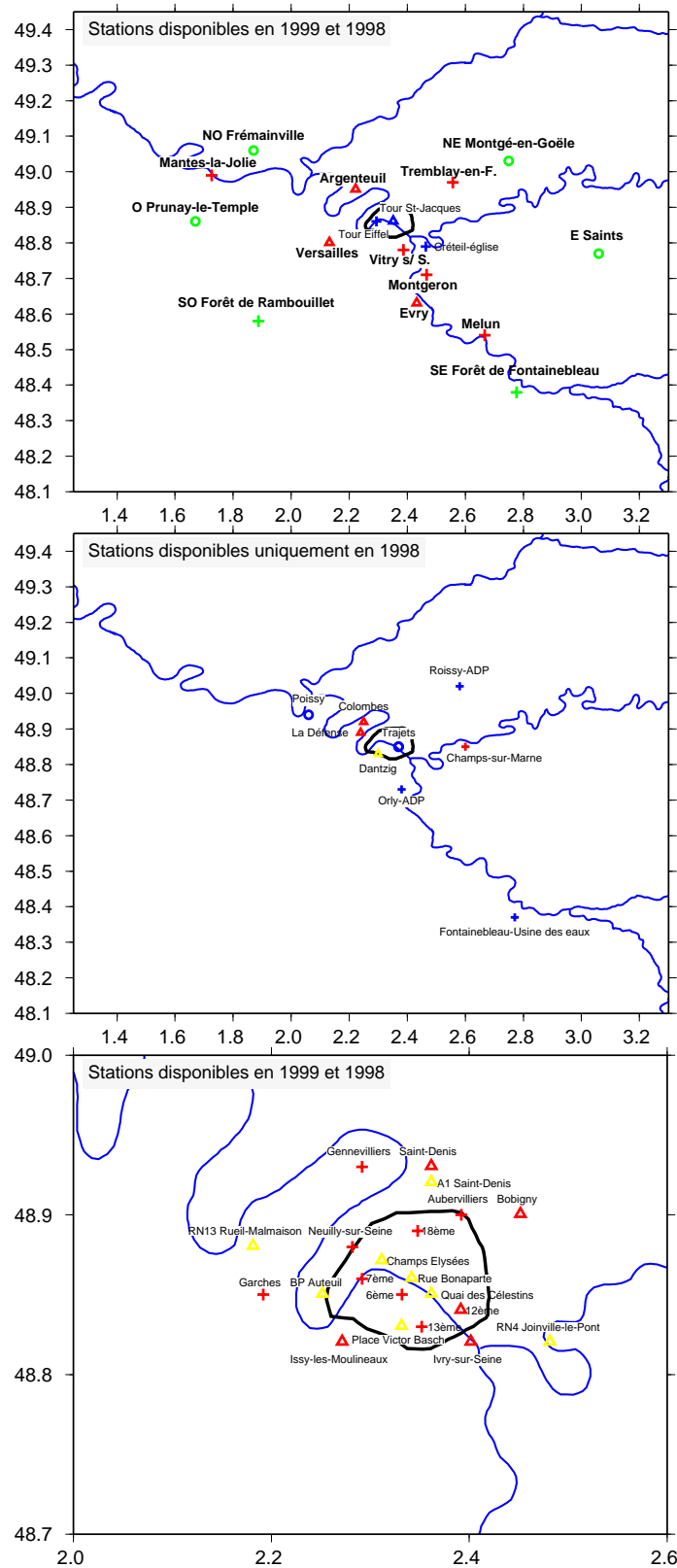


FIG. 6.2 – Stations de mesures AIRPARIF. Le type de station est indiqué en couleur : rouge = urbain, bleu = observation, vert = rural, jaune = trafic. Les mesures effectuées sont indiquées par le symbole : croix =  $O_3$ ,  $NO$  et  $NO_2$ , cercle =  $O_3$  seulement, triangle =  $NO$  et  $NO_2$  seulement.



pour l'ozone et  $\frac{1}{1,22}$  pour le monoxyde d'azote.

Enfin, lorsque plusieurs mesures sont disponibles pour une même maille, c'est leur moyenne qui est comparée à la valeur simulée.

**La simulation climatologique** Les champs de concentrations des étés 1998 et 1999 ont été simulés du 11 juin au 30 septembre, avec mise à l'équilibre du 1<sup>er</sup> au 10 juin avec le modèle CHIMERE version V200211K en mode prévision décrit au Chapitre 3. L'Île-de-France correspond, dans le modèle, à un domaine de  $150 \times 150$  km centré sur Paris. La résolution du domaine a été choisie pour coïncider avec celle du cadastre d'émissions, qui est de 6 km. Le maillage compte donc  $25 \times 25$  mailles horizontalement pour huit niveaux verticaux.

### 6.2.2.2 Les indicateurs statistiques

Dans tout ce qui suit, on note  $O_n$  la  $n$ -ième valeur observée de l'échantillon considéré de taille  $N$  et  $S_n$  la  $n$ -ième valeur simulée. La moyenne de chacune de ces variables est :

$$\overline{X_N} = \frac{1}{N} \sum_{i=1}^N X_n$$

Suivant (Blond 2002a), on étudie les performances du modèle avec les trois indicateurs statistiques suivants :

- **le biais** : c'est l'écart moyen entre les valeurs simulées et observées. Il indique les erreurs systématiques. On a :

$$bias = \frac{1}{N} \sum_{i=1}^N (S_n - O_n)$$

- **la moyenne quadratique** : elle donne l'erreur moyenne en valeur absolue. On a :

$$rms = \sqrt{\frac{1}{N} \sum_{i=1}^N (S_n - O_n)^2}$$

- **la corrélation** : elle permet de savoir si la variabilité est bien reproduite, indépendamment de l'amplitude. On a :

$$cor = \frac{\frac{1}{N} \sum_{i=1}^N (S_n - \overline{S_N}) \times (O_n - \overline{O_N})}{\sqrt{\frac{1}{N} \sum_{i=1}^N (S_n - \overline{S_N})^2 \times \frac{1}{N} \sum_{i=1}^N (O_n - \overline{O_N})^2}}$$

Ces variables permettent d'évaluer statistiquement la qualité du modèle si l'on suppose que les observations ne sont pas biaisées. Cette hypothèse est fautive : les mesures sont effectivement biaisées, leur contribution au biais étant due principalement aux erreurs de représentativité de la mesure par rapport à la maille du modèle (au sol, les erreurs de

mesure liées à l'instrument sont connues et très faibles). En réalité, on ne peut pas attribuer les parts d'erreur respectives dues à la mesure ou au modèle et celles-ci peuvent de plus se compenser. Ce n'est pas le but de cette étude puisque le krigeage permet d'obtenir le meilleur état des connaissances prenant en compte les mesures et les simulations. Il s'agit ici de détecter les principaux points faibles dans les paramètres du modèle, en considérant qu'un écart important avec les mesures en constitue un indicateur.

### 6.2.2.3 Résultats statistiques

#### L'ozone :

L'ozone est un polluant secondaire représentatif à la fois de toute la chimie troposphérique et des émissions. Les écarts entre concentrations d'ozone mesurées et simulées apportent donc des informations sur l'ensemble des processus en jeu.

<i>Statistiques ozone</i>	<i>été 1998</i>			<i>été 1999</i>		
<b>Site</b>	<b>biais</b>	<b>rms</b>	<b>cor</b>	<b>biais</b>	<b>rms</b>	<b>cor</b>
<b>zones urbaines</b>						
Champs-sur-Marne	7.5	13.3	0.76	x	x	x
Créteil	5.0	12.3	0.75	7.5	13.1	0.79
Garches	7.1	11.3	0.59	1.1	12.5	0.76
Gennevilliers	x	x	x	3.3	11.5	0.81
Mantes	6.5	12.4	0.79	3.2	11.6	0.81
Melun	x	x	x	7.6	13.1	0.81
Montgeron	2.9	11.5	0.78	6.3	12.3	0.80
Paris Nord-Est	0.8	10.2	0.8	0.5	11.1	0.8
Paris Nord-Ouest	4.6	11.0	0.8	4.6	11.8	0.79
Paris Sud-Est	-2.8	10.9	0.80	-1.3	11.8	0.78
Paris Sud-Ouest	1.1	10.7	0.78	0.2	11.4	0.78
Tremblay	-1.8	11.6	0.77	0.3	12.	0.78
Vitry	0.1	10.5	0.64	1.9	12.0	0.77
<b>zones rurales</b>						
zone rurale Est	x	x	x	5.1	11.1	0.81
zone rurale Nord-Est	0.2	10.9	0.75	-0.5	11.9	0.71
zone rurale Nord-Ouest	8.8	11.7	0.82	0.5	10.0	0.82
zone rurale Ouest	x	x	x	5.2	11.1	0.82
zone rurale Sud-Est	x	x	x	5.6	11.7	0.82
zone rurale Sud-Ouest	-0.7	9.5	0.83	-1.7	10.8	0.84

TAB. 6.1 – Valeurs des variables statistiques (biais, moyenne quadratique (rms) et corrélation (cor)) calculées sur tout l'été en 1998 et 1999 pour les concentrations d'ozone (en ppb) dans les différentes mailles où se trouvent des stations de mesure.

- **Le biais** dans l'estimation des concentrations d'ozone est compris entre -2,8 et 8,8 ppb (Tableau 6.1). Dans les zones rurales, où la représentativité des mesures est a priori meilleure que dans les zones urbaines, le biais est centré sur 0,5 ppb : le modèle a tendance à surestimer les concentrations d'ozone. Dans les zones urbaines, le biais est

centré sur 2,9 ppb ce qui indique une tendance à la surestimation. Ces tendances globales peuvent cependant résulter de la compensation de comportements très différents selon l'heure.

- Les **erreurs moyennes** sont comprises entre 9,5 et 13,3 ppb (Tableau 6.1), inférieures à 12 ppb dans les zones rurales. Le modèle étant particulièrement adapté à la reproduction des pics, ces erreurs restent assez faibles par rapport aux concentrations maximales atteintes (plus de 120 ppb pour une alerte).
- Enfin, la **corrélacion** est comprise entre 0,59 et 0,84 (Tableau 6.1) ce qui indique globalement une bonne reproduction des variations de concentration d'ozone. Cependant, étant donné les fortes variations de concentration d'ozone au cours d'une journée, la tendance globale peut dissimuler de fortes disparités selon les heures. La répartition géographique des corrélacions (Figure 6.3) montre que celles-ci sont plus élevées dans les zones rurales et les zones urbaines éloignées de Paris. Deux effets concourent à une répartition aussi nette : d'une part, la représentativité des stations rurales est meilleure et d'autre part, la corrélation de l'ozone en ville peut être abaissée par des défauts dans le profil temporel des émissions de  $\text{NO}_x$  qui titrent l'ozone rapidement.

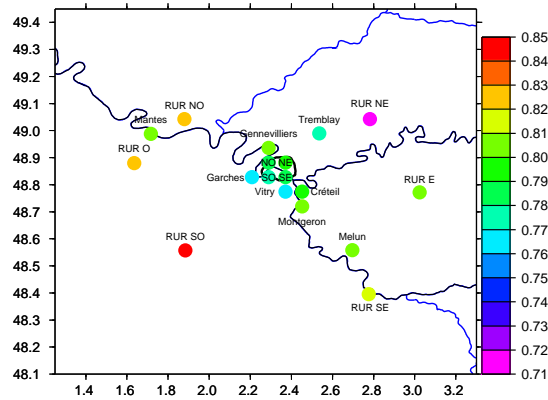


FIG. 6.3 – Répartition géographique de la corrélation calculée pendant l'été 1999 pour les concentrations d'ozone (ppb).

On s'intéresse donc plus précisément aux variations temporelles du biais et de la corrélation selon les heures de la journée. Parmi toutes les erreurs possibles, on notera que la représentativité des mesures peut varier en fonction du cas étudié : si l'on étudie un cas de forte pollution à l'ozone homogène sur tout le domaine, la mesure sera plus représentative de son environnement que si l'on étudie uniquement des cas de pollution locale avec des panaches très concentrés.

Les variations du biais dans l'estimation des concentrations d'ozone selon l'heure de la journée (Figure 6.4) indiquent, comme le confirme l'évolution de la corrélation (Figure 6.5), que le modèle ne s'attache pas particulièrement à reproduire les faibles concentrations nocturnes et leurs variations.

Pendant la journée, les concentrations d'ozone simulées l'après-midi sont bien estimées relativement à l'intensité d'un pic dans les régions rurales (entre -3,4 et +8,4 ppb à 15 heures) mais surestimées en zone urbaine (jusqu'à +9,7 ppb à 15 heures et +11,5 ppb à

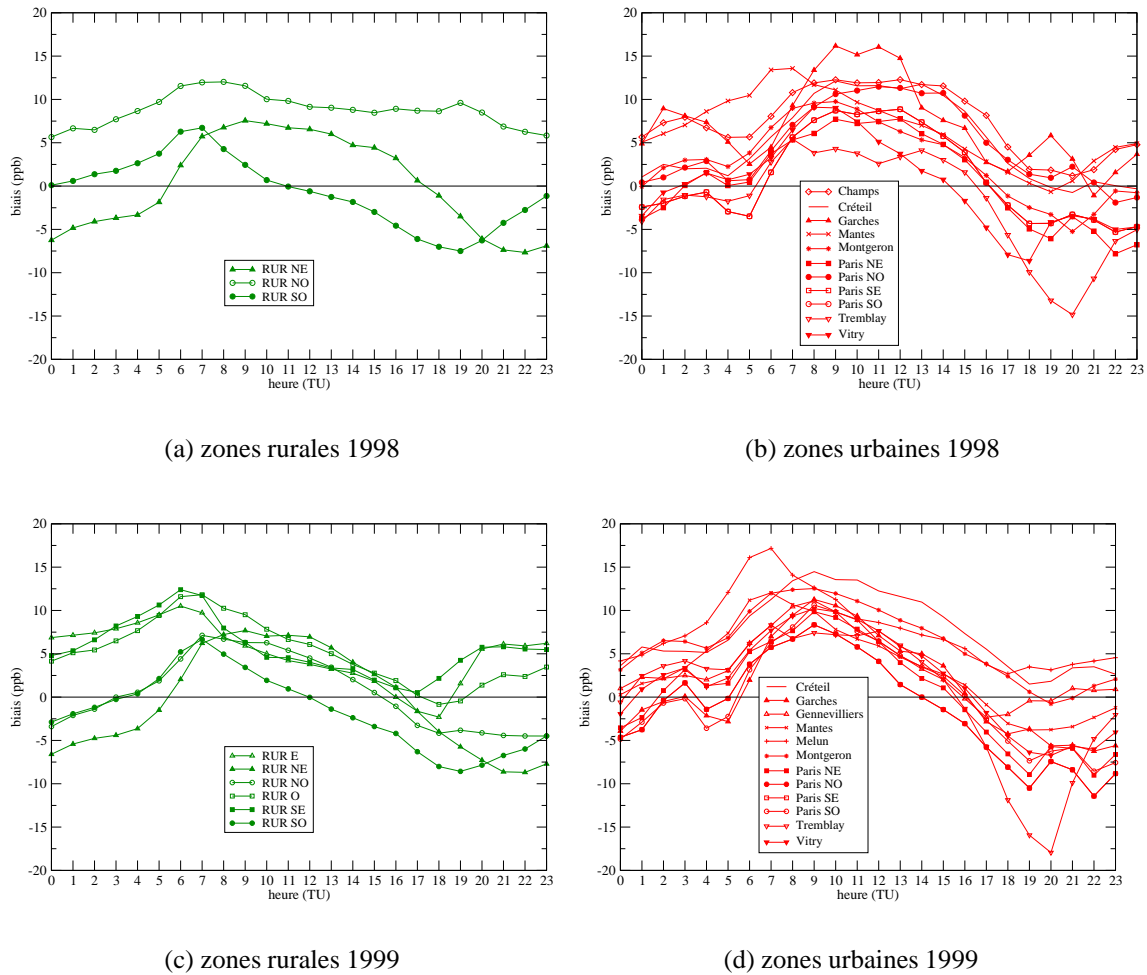


FIG. 6.4 – Évolution du biais selon l'heure de la journée pendant les étés 1998 et 1999 pour les concentrations d'ozone (ppb). En été, TU = heure légale - 2 heures.

14 heures) (Figure 6.4). Ces biais sont directement influencés par les données météorologiques d'entrée du modèle puisqu'un décalage de la position du panache de la ville par le modèle, qui se produit notamment quand le vent est faible, induit bien sûr un biais important avec les mesures (voir l'exemple page 81 avec la station de la zone rurale Sud-Ouest). L'évolution temporelle de la corrélation (Figure 6.5) indique, indépendamment des problèmes d'intensité, que le modèle reproduit bien l'évolution des concentrations d'ozone pendant la journée et particulièrement les pics de l'après-midi (corrélation comprise entre 0,78 et 0,93 à 15 heures). La zone rurale Sud-Ouest, qui est souvent située sous le vent de la ville et dans son panache d'ozone, présente les corrélations moyennes les plus grandes (0,83 et 0,84, Tableau 6.1).

*Cas particuliers des stations de Garches et Vitry en 1998 :* les corrélations obtenues dans ces deux stations sont négatives entre 8 heures et 10 heures à Garches (-0,3 à -0,4) et quasiment nulles à 10 et 11 heures à Vitry. Ces heures étant celles du pic de trafic du matin, ceci pourrait être dû à un problème de représentativité ou de modélisation du trafic routier. À Garches, les corrélations négatives correspondent à des biais élevés : 13,4, 16,2 et

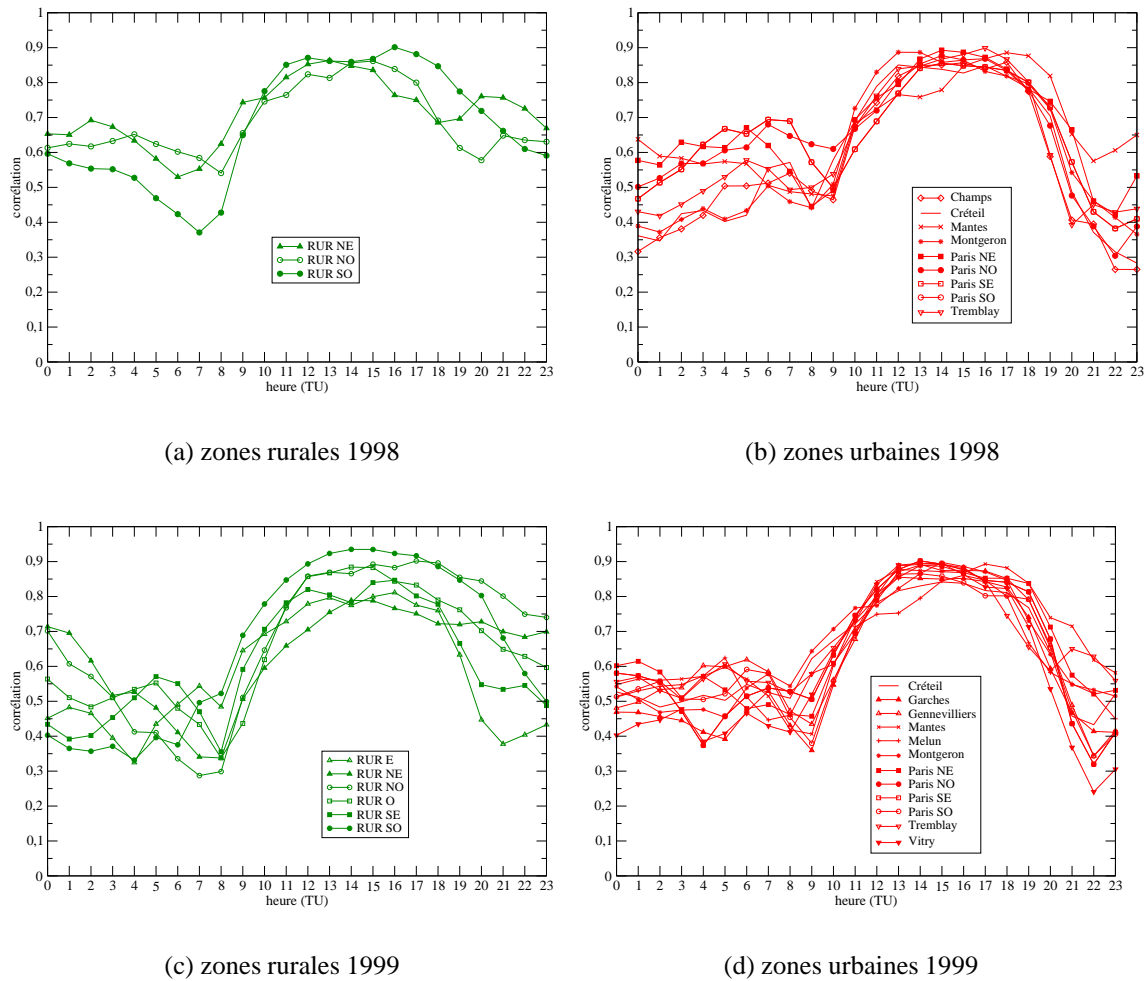


FIG. 6.5 – Évolution de la corrélation selon l'heure de la journée pendant les étés 1998 et 1999 pour les concentrations d'ozone (ppb). En été, TU = heure légale - 2 heures.

15,2 ppb. Pour les deux stations, les corrélations sont dans la moyenne des autres stations comparables le reste du temps et notamment l'après-midi. Ceci indique que le problème doit être lié au trafic. La titration de l'ozone autour de la station par les émissions dues au trafic routier du matin conduit à une baisse des concentrations d'après la mesure alors qu'à l'échelle de la maille du modèle, elles augmentent.

Finalement, l'étude des corrélations entre concentrations d'ozone mesurées et simulées permet d'affirmer que le **cycle diurne** de l'ozone est bien représenté dans le modèle. Le biais et la moyenne quadratique indiquent de plus que **l'intensité des pics** est bien reproduite malgré une certaine sensibilité aux **données météorologiques** comme la direction du vent qui peut induire un décalage du panache. Certains aspects particuliers, comme la répartition géographique des corrélations moyennes ou les cas des stations de Garches et Vitry, amènent à examiner les concentrations de monoxyde d'azote, leur profil temporel pouvant présenter des défauts dus aux émissions du trafic.

### Le monoxyde d'azote :

- **Le biais** dans l'estimation des concentrations de NO pour tout l'été est compris entre -9,9 et +0,2 ppb (Tableau 6.2) pour les zones urbaines et rurales. Les calculs sur les stations «trafic» ont été faits à partir d'une simulation comprenant une «couche d'émissions» c'est-à-dire un niveau supplémentaire représentant une fraction du premier des huit niveaux du domaine dans lequel les émissions sont injectées alors que la dilution n'intervient pas. Malgré cet artifice, les stations «trafic» ont des biais trop importants (jusqu'à -66 et +125 ppb) : elles ne sont pas assez représentatives par rapport à la maille du modèle.
- Cependant, **les corrélations** obtenues pour ces stations sont en général supérieures à celles des autres stations : 14 stations urbaines ou rurales sur 36 ont des corrélations supérieures ou égales à 0,6 contre 10 sur 13 pour les stations «trafic» (Tableau 6.2). Ceci pourrait indiquer une reproduction correcte du cycle journalier des émissions de  $\text{NO}_x$  dues au trafic. La répartition géographique des corrélations (Figure 6.6) montre en effet que celles-ci sont d'autant plus élevées que le lieu considéré est proche de Paris ou contient un grand axe routier.

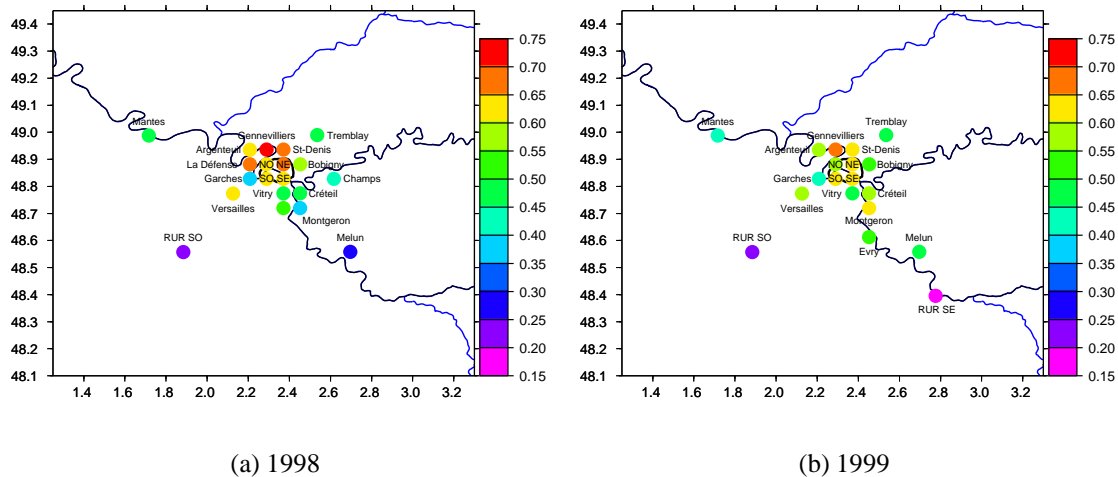


FIG. 6.6 – Répartition géographique de la corrélation calculée sur tout l'été en 1998 et 1999 pour les concentrations de NO (ppb).

- Enfin, **les erreurs moyennes** des stations urbaines sont comprises entre 6,4 et 23,7 ppb (Tableau 6.2) ; dans les deux zones rurales, l'erreur moyenne est inférieure à 4 ppb. Ces erreurs sont élevées par rapport aux concentrations maximales : les pics correspondant aux heures de pointe du trafic dépassent rarement 80 ppb en ville.

Les variations temporelles du biais dans l'estimation des concentrations de NO indiquent une tendance à la sous-estimation de plus de 10 ppb le matin dans les zones urbaines (Figure 6.7 (a) à (d)), particulièrement entre 6 et 9 heures, c'est-à-dire aux heures du pic matinal des émissions de  $\text{NO}_x$  dues au trafic. Dans les deux zones rurales disponibles, le biais est quasiment constant au cours de la journée avec une sous-estimation inférieure à 5 ppb (Figure 6.7 (b) et (d)).

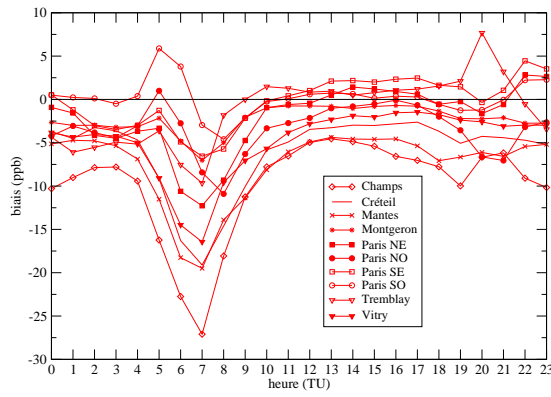
*Cas particulier de Garches en 1998, qui présente aussi un problème pour l'ozone* : les biais atteignent -39 à -87 ppb de 6 à 9 heures (non représenté). Le NO étant très largement

<i>Statistiques NO</i>	<i>été 1998</i>			<i>été 1999</i>		
<b>Site</b>	<b>biais</b>	<b>rms</b>	<b>cor</b>	<b>biais</b>	<b>rms</b>	<b>cor</b>
<b>zones urbaines</b>						
Argenteuil	-5.8	15.5	0.63	-5.4	13.3	0.56
Bobigny	-4.6	18.1	0.55	-5.1	16.9	0.54
Champs-sur-Marne	-9.9	21.	0.41	x	x	x
Créteil	-6.0	17.4	0.46	-5.9	16.3	0.55
Évry	x	x	x	-7.	16.2	0.50
Garches	-6.5	23.7	0.4	-0.1	9.3	0.44
Gennevilliers	-5.9	22.7	0.73	-4.7	18.9	0.67
La Défense	-7.4	18.	0.66	x	x	x
Mantes	-7.5	14.4	0.46	-3.4	7.3	0.45
Melun	-7.0	11.5	0.29	-3.3	6.4	0.46
Montgeron	-2.3	8.4	0.39	-3.2	7.4	0.62
Paris Nord-Est	-2.1	14.2	0.69	-2.5	14.5	0.57
Paris Nord-Ouest	-3.4	14.9	0.60	-2.8	13.5	0.56
Paris Sud-Est	-0.3	14.4	0.62	-0.8	12.7	0.61
Paris Sud-Ouest	0.2	10.	0.65	-1.3	10.3	0.60
Saint-Denis	-5.5	19.3	0.66	-3.9	17.2	0.64
Tremblay	-1.3	11.0	0.46	-2.5	14.2	0.48
Versailles	-3.3	9.6	0.61	-2.7	8.4	0.56
Vitry	-4.8	18.6	0.5	-5.1	18.3	0.48
<b>zones rurales</b>						
zone rurale Sud-Est	x	x	x	-2.1	2.4	0.16
zone rurale Sud-Ouest	-2.1	3.3	0.25	-1.9	2.7	0.21
<b>stations trafic : avec couche d'émissions</b>						
A1 Saint-Denis	-66.4	104.3	0.68	-41.5	81.3	0.65
Basch+BP Auteuil+Dantzig	7.1	52.3	0.67	-42.9	75.6	0.71
Rue Bonaparte	118.4	125.9	0.7	125.3	134.2	0.66
Champs-Élysées	107.8	122.	0.51	118.3	131.4	0.48
Quai des Célestins	60.2	93.7	0.64	45.4	90.8	0.63
RN4 Joinville	27.3	59.3	0.6	21.2	59.6	0.62
RN13 Rueil-Malmaison	x	x	x	34.7	51.8	0.65

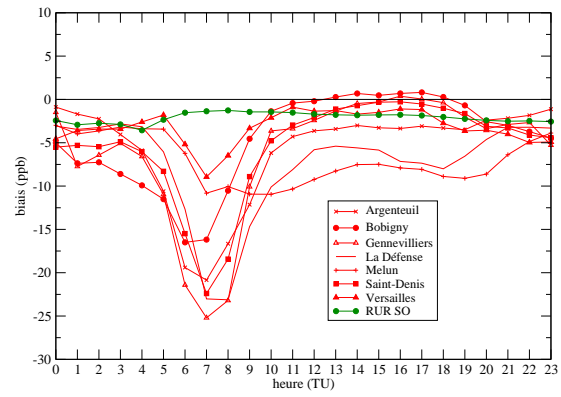
TAB. 6.2 – Valeurs des variables statistiques (biais, moyenne quadratique (rms) et corrélation (cor)) calculées sur tout l'été en 1998 et 1999 pour les concentrations de NO (en ppb) dans les différentes mailles où se trouvent des stations de mesure.

sous-estimé à cet endroit à ces heures particulières, il s'agit donc très probablement d'un problème lié au trafic. L'ozone est titré près de la station mais pas dans le modèle, pour lequel le NO est plus dilué dans cette maille. Il y a donc un problème de représentativité de la mesure (qui disparaît en 1999, les stations AIRPARIF ayant toutes été revues et modifiées).

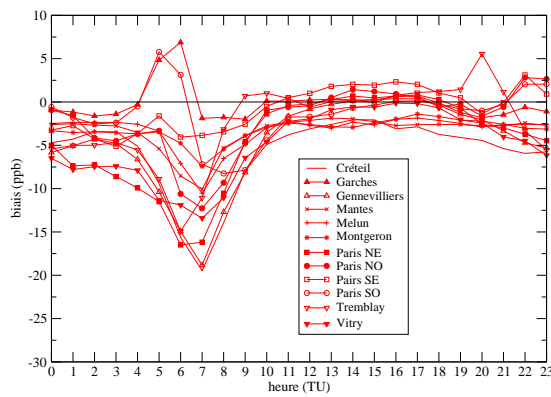
Finalement, l'étude des corrélations entre concentrations de monoxyde d'azote mesurées et simulées indique que la **variabilité et le phasage temporels** du NO en milieu urbain



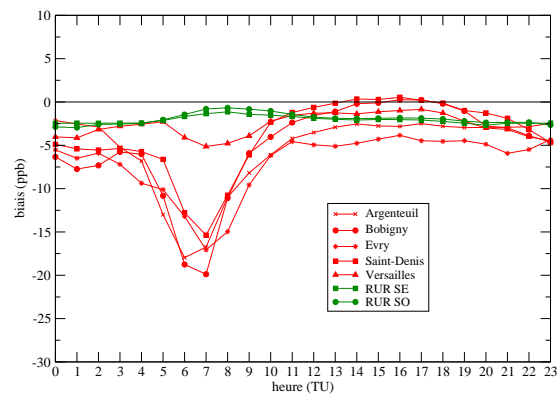
(a) mêmes zones urbaines que l'ozone 1998



(b) zones urbaines supplémentaires et zones rurales 1998



(c) mêmes zones urbaines que l'ozone 1999



(d) zones urbaines supplémentaires et zones rurales 1999

FIG. 6.7 – Évolution du biais selon l'heure de la journée pendant l'été 1999 pour les concentrations de NO (ppb). En été, TU = heure légale - 2 heures.

sont correctement représentés. Cependant, le biais et la moyenne quadratique indiquent que l'intensité en est mal reproduite, probablement à cause de défauts dans **l'amplitude** des flux émis par le trafic. On peut également évoquer l'hypothèse d'une imperfection dans la paramétrisation du **mélange vertical**. Un déclenchement trop rapide de celui-ci pourrait conduire à sous-estimer les concentrations de NO matinales. L'évolution temporelle des corrélations (non représentée) indique cependant que l'heure de déclenchement est correcte. Il reste donc à examiner l'intensité du mélange vertical, qui dépend principalement du coefficient de mélange turbulent et de la hauteur de couche-limite (Vautard et al. 2003). L'influence de ces deux variables sera donc étudiée dans ce qui suit avant de procéder à toute inversion.



## 6.3 Choix des cas à étudier

### 6.3.1 Hypothèses

#### 6.3.1.1 Le modèle

On considère dans ce qui suit qu'aucun processus ne manque dans le modèle (voir page 85) et que toutes les paramétrisations nécessaires sont suffisamment réalistes. De plus, lorsqu'il n'y pas de correction préalable ou d'inversion de paramètres météorologiques, cela signifie que les conditions aux limites et la météorologie, notamment la vitesse et la direction du vent et la température, sont considérées comme assez précises pour que les défauts de la simulation directe ne leur soient pas majoritairement imputables. Ces hypothèses de travail ne sont pas totalement exactes mais elles sont nécessaires et assez réalistes, le modèle simulant déjà de façon satisfaisante la structure des champs de concentrations aux échelles continentale et régionale (ESQUIF 2001; Blond 2002b).

**Le critère d'arrêt :** la valeur théorique estimée page 51 est trop petite pour être applicable à des cas réels. On considère donc que la différence entre deux termes en ppb est non significative quand elle est inférieure à non pas  $10^{-2}$  mais seulement  $10^{-1}$ , ce qui correspond à  $eps_g = \frac{\sqrt{dim} \times 10^{-1} \times \sum_{i=1}^N w_i}{\|g_1\|}$ .

#### 6.3.1.2 Le cadastre

**Émissions bien placées** Comme indiqué précédemment (voir page 35), il peut être impossible de distinguer entre un mauvais placement de la source et une amplitude mal estimée. Dans toute notre étude, on suppose que les sources d'émissions sont bien placées et que seule leur intensité doit être modifiée. Cette hypothèse est réaliste, les nombreuses études menées en Île-de-France ayant permis de corriger les éventuelles grosses erreurs de positionnement (Vautard et al. 2003).

**Familles d'espèces** L'inversion des familles ( $\text{NO}_x$ , COV) est faite en supposant que l'agrégation des espèces émises au sein de chaque famille est suffisamment réaliste. Cela signifie que les rapports fournis par le cadastre *a priori* entre les différentes espèces composant la famille, par exemple le NO et le  $\text{NO}_2$ , sont conservés lors de l'inversion ; c'est donc l'intensité du flux de toute la famille qui est inversée.

### 6.3.2 Le problème à inverser

#### 6.3.2.1 Émissions à optimiser

Les résultats statistiques obtenus pour les concentrations dans la Section précédente suggèrent que les émissions du pic matinal de  $\text{NO}_x$  en Île-de-France pourraient être mal évaluées dans le cadastre *a priori*. Le rôle du mélange vertical dans les écarts entre concentrations mesurées et simulées doit cependant être étudié avant de procéder éventuellement à l'inversion de ces émissions. Dans ce cas, il apparaît donc nécessaire d'inverser les émissions de  $\text{NO}_x$  afin d'obtenir un profil temporel optimisé pour le pic de concentration du matin, de 4 à 11 heures d'après les évolutions du biais (voir page 96).

Le problème concernant les flux en milieu urbain (voir page 95), seuls sont optimisés les flux les plus intenses c'est-à-dire ceux qui recouvrent les zones de trafic routier. Cela correspond aux flux de  $\text{NO}_x$  supérieurs à  $5 \times 10^{11} \text{moléc.cm}^{-2}.\text{s}^{-1}$ , soit environ 30 à 60 flux pour chacune des huit heures de la fenêtre temporelle définie précédemment. La Figure 6.8 représente les émissions de NO (soit 90% des émissions de  $\text{NO}_x$ ) *a priori* à 8 heures : les flux non inversés sont en blanc et correspondent aux environnements ruraux ou péri-urbains peu denses. L'agrégation spatiale est effectuée sur les flux sélectionnés comme

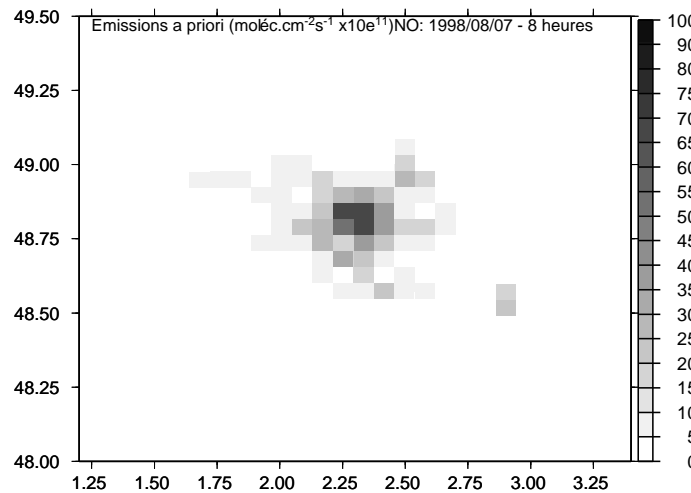


FIG. 6.8 – Cadastre d'émissions *a priori* : flux de NO le 7 août 1998 à 8 heures TU. En été, TU = heure légale - 2 heures.

décrit page 72. Cette technique étant adaptative, le nombre et la taille des zones définies varient à chaque heure.

### 6.3.2.2 Contraintes

Les contraintes utilisées pour l'inversion des émissions de  $\text{NO}_x$  sont les analyses de NO effectuées de 4 à 16 heures. En effet, les résultats obtenus au Chapitre 4 (page 58) indiquent que les concentrations d'ozone mesurées en Île-de-France ne contiennent pas assez d'information sur les émissions de  $\text{NO}_x$  matinales. Le krigeage des concentrations est effectué grâce aux mesures du réseau AIRPARIF décrites page 86. Les stations «observation», dont la représentativité n'est pas contrôlée, et les stations «trafic», dont la représentativité n'est pas adaptée au problème (trop petite par rapport à la maille du modèle), ne sont pas utilisées pour l'analyse.

Dans un premier temps seront inversés des cas particuliers pour lesquels de nombreuses données et études complémentaires sont disponibles. Ces études devraient permettre de déterminer à quelles conditions il est possible d'appliquer la méthode d'optimisation à d'autres périodes.

## 6.4 Le 7 août 1998

### 6.4.1 Choix de cette journée

Le vendredi 7 août 1998 est le deuxième jour de la POI2 (Période d'Observation Intensive) de la campagne ESQUIF. Pendant cette journée, le vent est resté faible (inférieur à  $3\text{m.s}^{-1}$ ), les températures élevées (caniculaires avec des relevés dépassant  $38^{\circ}\text{C}$ ) et les masses d'air importées dans le domaine étaient peu chargées en polluants car d'origine océanique (ESQUIF 2001; Menut et al. 2000). Cette situation météorologique de stagnation couplée à une hauteur de couche-limite faible (moins de 1000 mètres à son maximum d'après les données de l'ECMWF, moins de 1300 mètres d'après les mesures ESQUIF, voir ci-dessous page 102) a conduit à un épisode de pollution dite «locale» lors duquel l'influence des émissions locales est prépondérante, par rapport notamment à celle du contenu des masses d'air advectées (Menut et al. 2000).

Un panache d'ozone s'est formé l'après-midi au Sud-Ouest de Paris, comme représenté sur la Figure 6.9. La production d'ozone est maximale dans le panache, à environ 50 km sous le vent de la ville elle-même. Des concentrations d'ozone atteignant 110 ppb ont été observées en fin d'après-midi alors que les concentrations au vent étaient de 60 à 70 ppb. La différence de concentration d'ozone due aux émissions locales était donc d'au moins 50 ppb ce jour-là. À part un léger décalage du panache, détaillé page 81, les mesures et le modèle sont globalement en bon accord (ESQUIF 2001; Hodzic 2002; Menut et al. 2000).

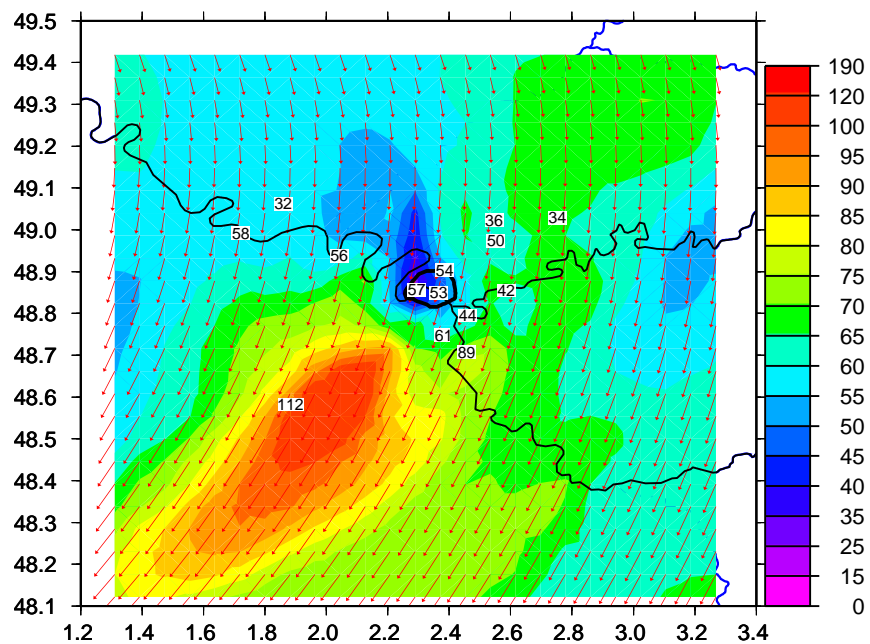


FIG. 6.9 – Concentrations d' $\text{O}_3$  (en ppb) en surface le 7 août à 15 heures TU. Les flèches rouges indiquent le sens et l'intensité du vent, les étiquettes correspondent aux valeurs relevées par des stations de mesure AIRPARIF. En été, TU = heure légale - 2 heures.

Cependant, une comparaison plus détaillée des concentrations d'ozone et de monoxyde d'azote mesurées et simulées indique que les concentrations de NO simulées le matin en

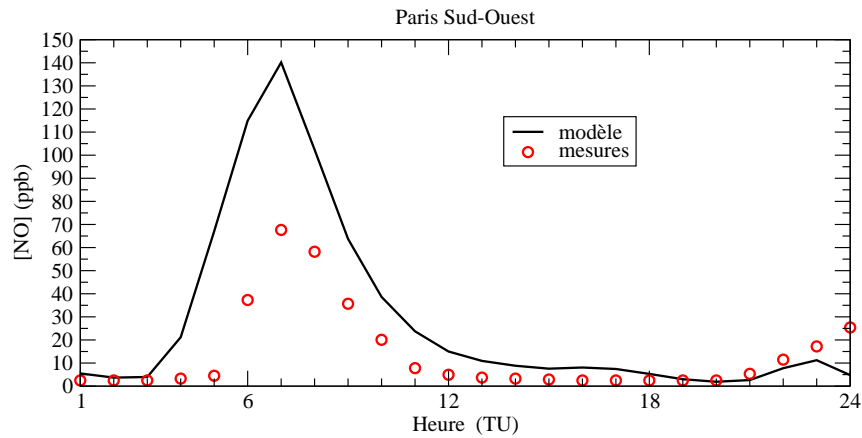


FIG. 6.10 – Comparaison entre les concentrations de monoxyde d’azote mesurées par le réseau AIRPRIF et calculées par le modèle pour le 7 août 1998. En été, TU = heure légale - 2 heures.

ville sont très supérieures aux concentrations mesurées : la Figure 6.10 représente la situation dans le quart Sud-Ouest de Paris, les concentrations simulées entre 6 et 10 heures étant 1,8 à 3,1 fois supérieures aux concentrations mesurées. Alors que les concentrations d’ozone mesurées et simulées sont en bon accord en amont de la ville, comme représenté pour la zone rurale Nord-Est sur la Figure 6.11, les concentrations simulées l’après-midi dans le panache sont sous-estimées, indépendamment du problème de centrage géographique de celui-ci : dans la zone rurale Sud-Ouest, les concentrations mesurées entre 15 et 20 heures sont supérieures de 8 à 219% aux concentrations simulées (Figure 6.11). Selon le régime chimique dominant dans le panache, la surestimation des concentrations de monoxyde d’azote en ville pourrait expliquer la sous-estimation des concentrations d’ozone l’après-midi dans le panache.

Comme indiqué page 97, la surestimation des concentrations simulées de NO (par rapport aux mesures) peut être due à une sous-estimation du mélange vertical ou à une surestimation des émissions de  $\text{NO}_x$ . L’influence du mélange vertical sur les concentrations simulées le 7 août 1998 sera donc estimée en examinant les rôles du coefficient de mélange turbulent et de la hauteur de couche-limite avant l’optimisation des émissions de  $\text{NO}_x$ .

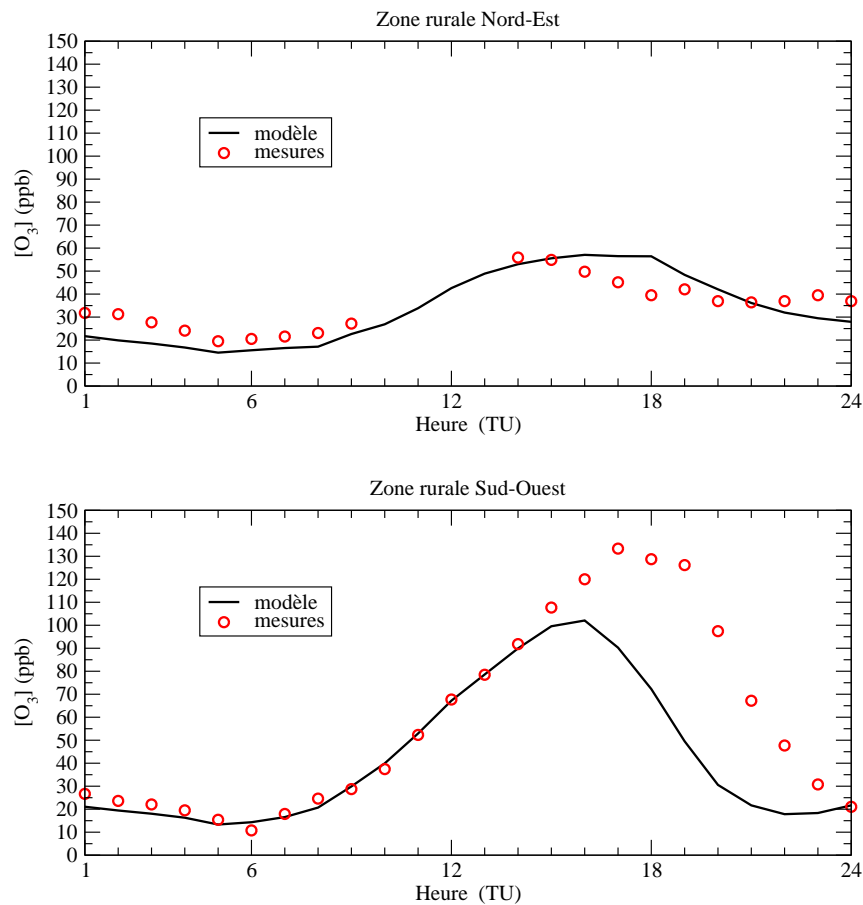


FIG. 6.11 – Comparaison entre les concentrations d’ozone mesurées par des stations AIRPARIF et calculées par le modèle le 7 août 1998. La Zone Rurale Nord-Est correspond à la station de Montgé-en-Goële (au vent) et la Zone Rurale Sud-Ouest, à la station de la forêt de Rambouillet (sous le vent) - voir Figure 6.2 page 88 pour l’emplacement des stations. En été, TU = heure légale - 2 heures.

## 6.4.2 Impact du mélange vertical

### 6.4.2.1 Coefficient de mélange turbulent

Pour estimer l'influence du mélange vertical, des simulations directes ont été effectuées en perturbant le coefficient de mélange turbulent  $K_z$ , pour tout le profil vertical et dans tout le domaine, de  $\pm 30\%$  entre 4 et 11 heures.

Par rapport à la simulation de référence, les concentrations de monoxyde d'azote simulées le matin en ville (Figure 6.12(a) en haut) sont augmentées au maximum de 16,4% lorsque le coefficient  $K_z$  est diminué de 30% et diminuées au maximum de 10,4% lorsque le coefficient  $K_z$  est augmenté de 30% (Figure 6.12(a) en bas). Ces perturbations n'ont pratiquement pas d'impact sur les concentrations d'ozone dans le panache l'après-midi, comme représenté sur la Figure 6.12(b) pour la zone rurale Sud-Ouest située dans le panache entre 12 et 17 heures.

Cette étude d'impact très simple indique que la paramétrisation du coefficient de mélange turbulent n'a pas une influence suffisante pour expliquer, à elle seule, les écarts entre mesures et modèle. Cependant, la hauteur de couche-limite, qui intervient dans la paramétrisation de  $K_z$ , est un paramètre comportant une grande incertitude.

### 6.4.2.2 Hauteur de couche-limite

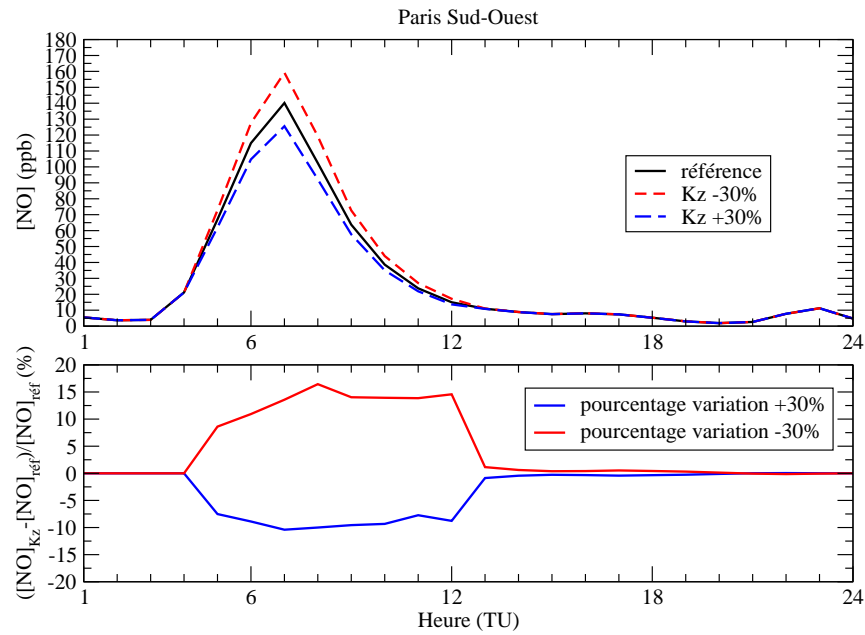
Pour la journée du 7 août 1998, faute de mesures par lidar, on ne dispose que des mesures obtenues par les sondages effectués à Trappes (base de données accessible en ligne sur <http://climserv.ipsl.polytechnique.fr/esquif/>), dont les résultats sont indiqués dans le Tableau 6.3. Comme les mesures plus nombreuses des 8 et 9 août (avec

Heure	HCL mesurée (m)	HCL modèle (m)	écart (%) $\frac{HCL_{mod} - HCL_{mes}}{HCL_{mes}}$
5,5	93	61	-34,4
8,45	229	148	-35,4
11,3	690	507	-26,5
14,5	923	734	-20,5
17,45	1267	535	-57,8

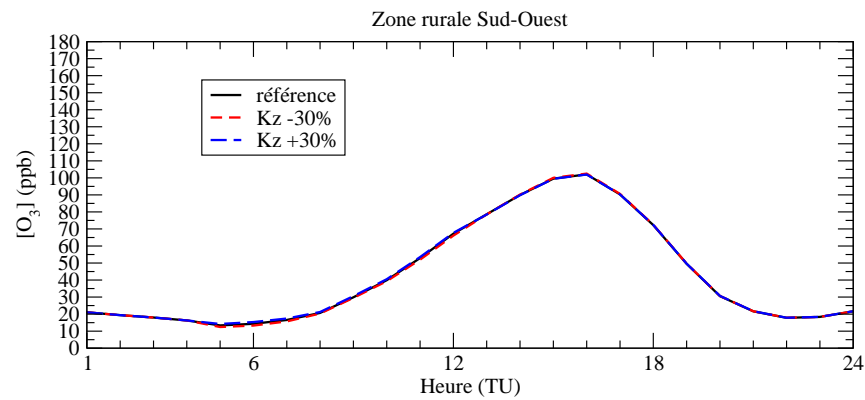
TAB. 6.3 – *Hauteur de couche-limite (m) mesurée et utilisée par le modèle le 7 août 1998.*

notamment le lidar de Palaiseau), les mesures du 7 août sont supérieures aux valeurs utilisées par la modèle de 20,5 à 57,8%. D'après les deux mesures disponibles à 5 heures 30 et à 8 heures 27, la HCL matinale utilisée par le modèle est sous-estimée d'environ 35% par rapport aux mesures.

Pour quantifier l'influence d'une éventuelle sous-estimation de la hauteur de couche-limite, deux simulations directes sont effectuées en augmentant la HCL de 50% pendant toute la journée afin de garder un profil temporel cohérent et sur tout le domaine afin de conserver le profil spatial dû à l'îlot de chaleur urbain (bien qu'il soit peu important en été). Pour la seconde simulation, le coefficient de mélange turbulent est de plus augmenté de 30% le matin.



(a) NO



(b) Ozone

FIG. 6.12 – Concentrations de monoxyde d’azote et d’ozone (ppb) obtenues avec différentes valeurs du coefficient de mélange turbulent  $K_z$ . En été, TU = heure légale - 2 heures.

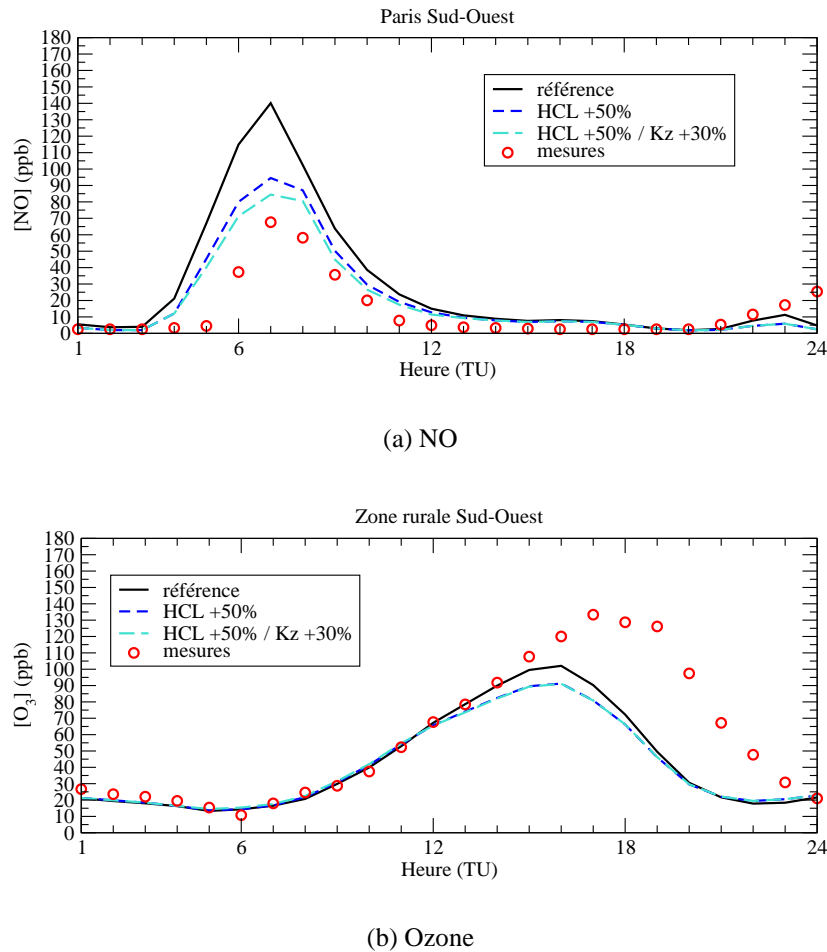


FIG. 6.13 – Concentrations de monoxyde d’azote et d’ozone (ppb) obtenues avec différentes HCL et valeurs de  $K_z$ . En été, TU = heure légale - 2 heures.

Les concentrations de NO simulées le matin en ville avec une HCL augmentée de moitié sont de 15 à 39,5% inférieures aux concentrations obtenues avec la HCL de référence, l’impact du  $K_z$  restant faible (Figure 6.13(a)) et sont nettement plus proches des valeurs mesurées. De plus, la sous-estimation des concentrations d’ozone simulées par rapport aux mesures est peu accentuée par l’augmentation de la HCL (Figure 6.13(b)).

### 6.4.2.3 Conclusion

La paramétrisation du coefficient de mélange turbulent comporte des bases physiques validées. Il paraît donc difficile de modifier arbitrairement le coefficient obtenu avec comme seul objectif de mieux reproduire les concentrations de monoxyde d’azote en certains lieux et à certaines heures. De plus, ces modifications du coefficient de mélange turbulent n’ont quasiment pas d’influence sur les écarts entre concentrations d’ozone simulées et mesurées dans le panache l’après-midi. On considère donc dans toute cette étude que la paramétrisation du coefficient de mélange turbulent  $K_z$  est correcte. Par contre, il est nécessaire, lorsque cela est possible, de débiaiser la hauteur de couche-limite avant de



procéder à l'inversion. Dans les inversions effectuées pour le 7 août 1998, la HCL est augmentée de 50% sur tout le domaine pendant vingt-quatre heures.

### 6.4.3 Optimisation des émissions

Pour l'optimisation effectuée ici, seuls les flux de  $\text{NO}_x$  sont inversés : le rapport COV sur  $\text{NO}_x$  à l'émission est donc modifié. L'agrégation spatiale est effectuée avec une tolérance  $\tau_E \geq 25\%$  qui permet de regrouper, pour la plupart des heures inversées, les quatre flux correspondant à la ville au sein d'une même zone. Deux cycles krigeage-optimisation sont enchaînés, de façon à améliorer la qualité des résultats par rapport à un seul cycle mais sans utiliser un temps de calcul trop long (voir page 66).

#### 6.4.3.1 Inversion des flux de $\text{NO}_x$

**Moyenne sur tout le domaine** La tendance globale donnée par l'optimisation aux flux d'émission de  $\text{NO}_x$  est indiquée par le profil temporel moyen sur tout le domaine (Figure 6.14). Par rapport au cadastre *a priori*, le pic du matin du cadastre optimisé est intensifié entre 4 et 5 heures et réduit les heures suivantes. L'augmentation la plus forte est obtenue à 5 heures avec un flux optimisé moyen plus intense de 11,7% que le flux *a priori* ; la réduction la plus forte est obtenue à 6 heures avec un flux optimisé moyen plus faible de 16,1% que le flux *a priori*. La masse moyenne émise au cours de la journée dans tout le domaine n'est donc diminuée que de 1% par l'optimisation. Ces corrections font apparaître sur le profil temporel moyen deux pics nettement séparés le matin, entre 4 et 5 heures et 8 et 9 heures, au lieu d'un plateau de 4 à 9 heures.

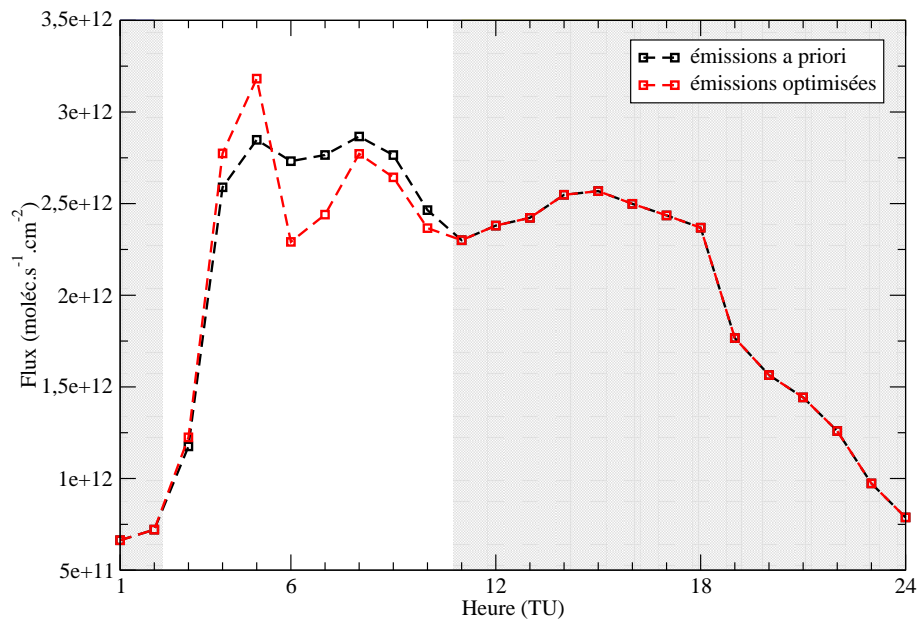
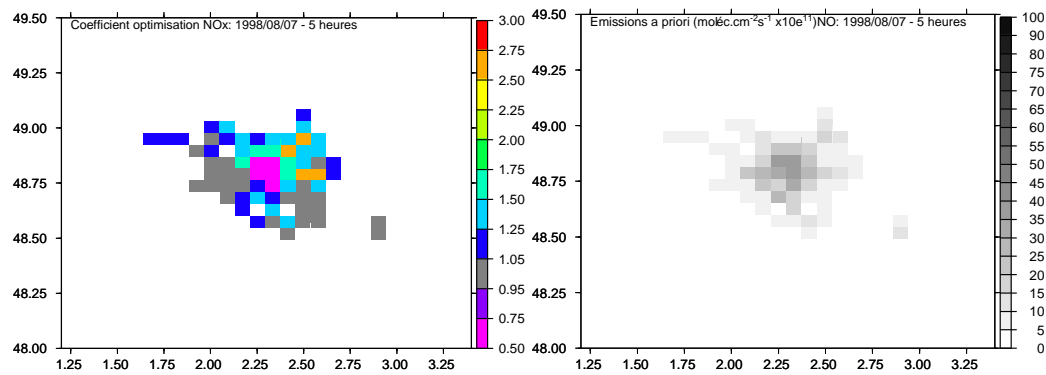


FIG. 6.14 – Profil temporel moyen sur tout le domaine du flux d'émission de  $\text{NO}$  le 7 août 1998. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.

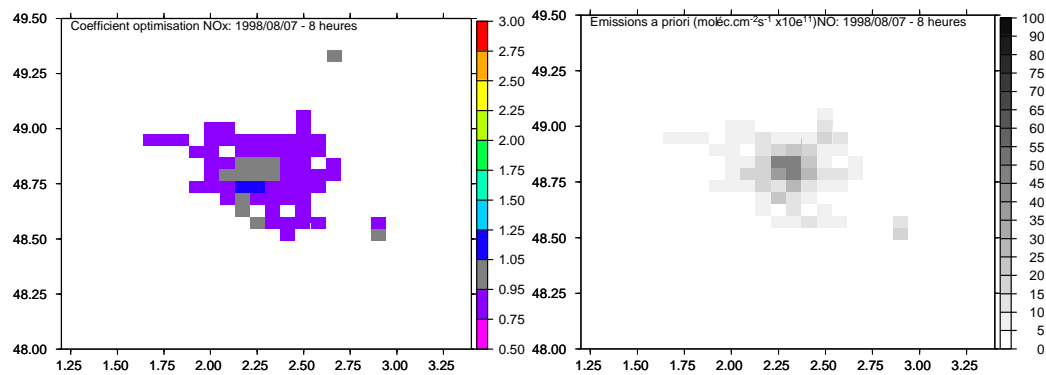
**Répartition spatiale** Afin de déterminer quelles émissions optimisées contribuent à ces deux pics différents, les répartitions spatiales des corrections à 5 et 8 heures sont détaillées sur la Figure 6.15. Celle-ci indique qu'à une heure donnée, toutes les corrections ne sont pas effectuées dans le même sens.

À 5 heures, les flux les plus intenses correspondant à la ville sont multipliés par 0,5 à 0,75 alors que les flux intenses extérieurs à Paris situés en amont de la ville sont multipliés par 1,25 à 2,75. Les flux les plus faibles sont peu modifiés, le coefficient de l'optimisation restant compris entre 0,95 et 1,25. Ainsi le premier pic des émissions optimisées moyennes sur tout le domaine est-il dû à la forte augmentation, après optimisation, des flux *a priori* assez intenses de lieux situés en amont de Paris elle-même.

À 8 heures, les flux les plus intenses sont modifiés de moins de 5% avec des coefficients de correction compris entre 0,95 et 1,05. Les flux moins intenses de l'extérieur de la ville sont réduits, multipliés par 0,75 à 0,95. Les corrections sont beaucoup plus homogènes qu'à 5 heures : le deuxième pic du profil temporel des émissions optimisées moyennes sur tout le domaine correspond donc à un abaissement général des flux d'émissions inversés.



(a) à 5 heures TU

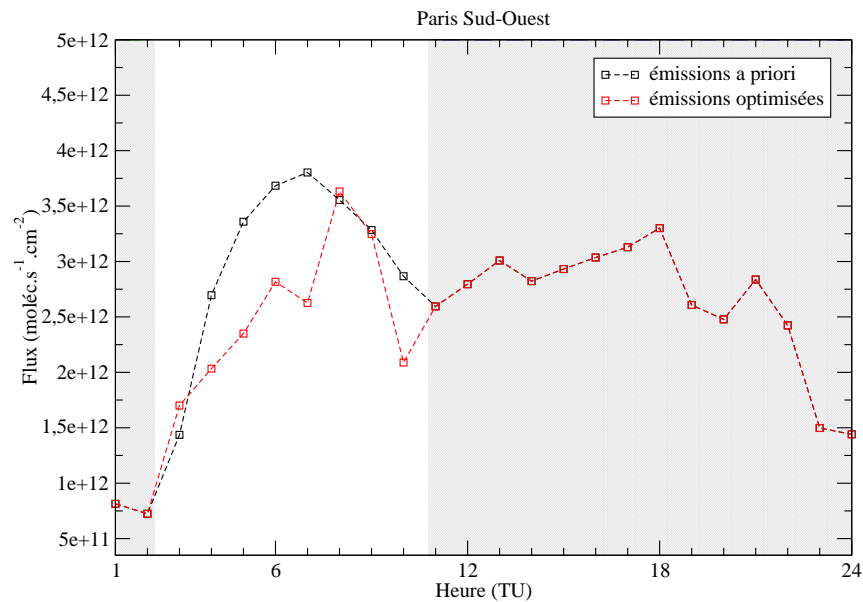


(b) à 8 heures TU

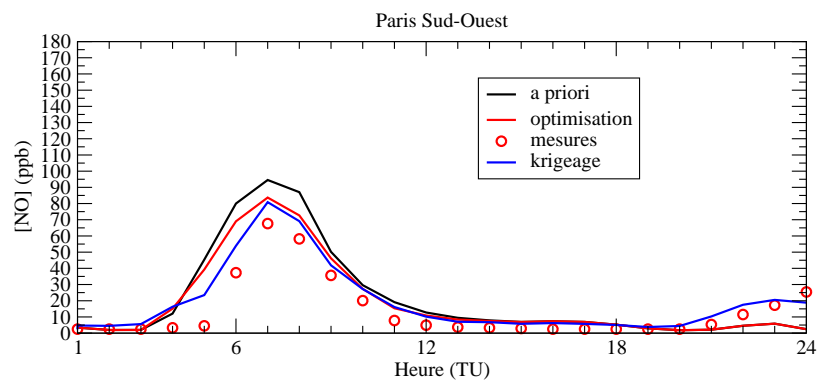
FIG. 6.15 – Répartition spatiale des corrections pour les  $\text{NO}_x$  et intensité des émissions de NO (représentant 90% des flux de  $\text{NO}_x$ ) le 7 août 1998. En été, TU = heure légale - 2 heures.

**Différents lieux** Le profil temporel moyen des émissions recouvre donc des situations différentes selon les lieux : dans Paris, en amont de la ville et aux endroits où les flux sont plus faibles.

L'évolution temporelle des flux d'émissions et des concentrations de monoxyde d'azote dans la ville sont représentés sur la Figure 6.16 pour la maille correspondant à Paris Sud-Ouest (représentative de la ville). La durée du pic d'émission matinal (flux supérieurs à  $3.10^{12} \text{moléc.cm}^{-2}.\text{s}^{-1}$  pour cette maille) est réduite de cinq heures, de 5 à 9 heures du matin, à deux heures, de 8 à 9 heures. L'effet général sur le profil temporel d'émissions



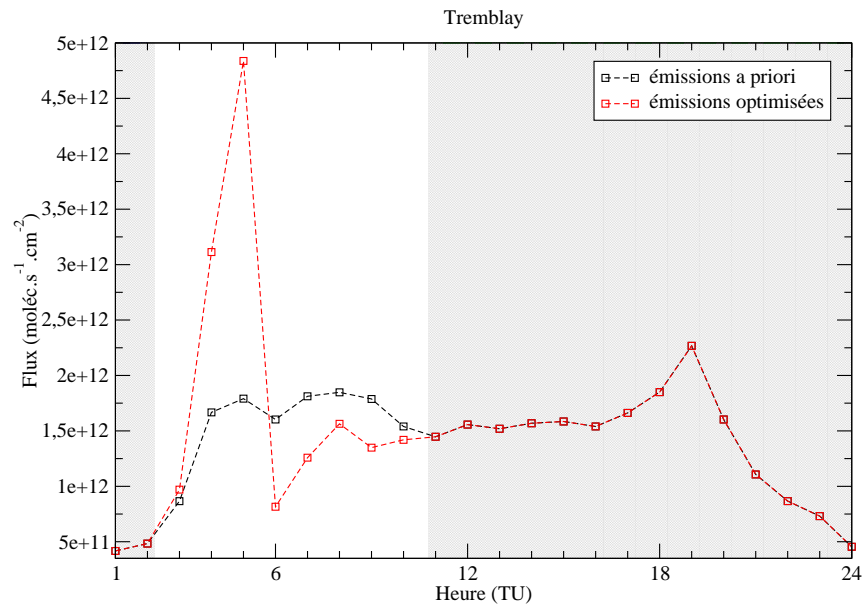
(a) Émissions a priori et optimisées



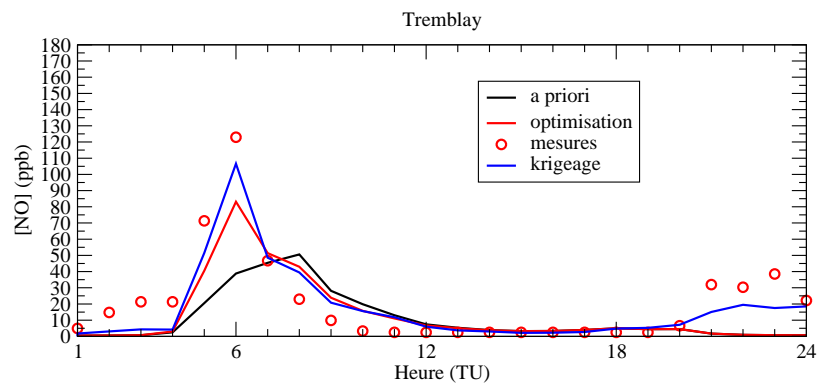
(b) Concentrations de monoxyde d'azote a priori, optimisées et mesurées

FIG. 6.16 – Évolutions temporelles des émissions et des concentrations de monoxyde d'azote le 7 août 1998 dans la maille Paris Sud-Ouest. Les concentrations d'une heure donnée résultent des émissions de l'heure précédente. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.

correspond au deuxième pic nettement séparé observé sur le profil moyen du cadastre optimisé (Figure 6.14). Les évolutions temporelles des concentrations de NO *a priori* et mesurées présentaient des allures semblables avec un pic matinal survenant à 7 heures. Les concentrations optimisées conservent le même profil temporel, seule leur intensité est modifiée, le pic passant de 94 ppb simulés *a priori* à 84 ppb pour une mesure de 68 ppb. À ce stade, le réalisme des émissions optimisées ne peut être estimé que sur des critères généraux : le profil temporel ne comporte pas d'oscillation ; le cadastre *a priori* ayant



(a) Émissions a priori et optimisées



(b) Concentrations de monoxyde d'azote a priori, optimisées et mesurées

FIG. 6.17 – Évolutions temporelles des émissions et des concentrations de monoxyde d'azote le 7 août 1998 dans la maille de Tremblay. Les concentrations d'une heure donnée résultent des émissions de l'heure précédente. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.

été construit sur la base du mois de juillet 1998, on peut envisager que le trafic moins dense du mois d'août aboutisse au retardement du pic en ville. En effet, les variations diurnes utilisées pour élaborer le cadastre *a priori* sont les mêmes en juillet et en août, seule l'intensité est à toute heure inférieure de 20% au mois d'août.

Dans quatre mailles situées en amont de la ville (représentées en orange sur la Figure 6.15 (a)), les flux d'émissions ont été fortement corrigés pendant une ou deux heures. L'évolution temporelle des flux d'émissions et des concentrations de monoxyde d'azote dans la maille de Tremblay, qui en fait partie, sont représentés sur la Figure 6.17. Le pic de concentrations de NO mesuré était plus matinal et plus intense que le pic simulé *a priori* : 123 ppb à 6 heures contre 51 ppb à 8 heures. Pour approcher les contraintes en intensité et en temps, les émissions de NO<sub>x</sub> *a priori* sont multipliées par 1,9 à 4 heures et 2,7 à 5 heures. Cette modification importante des intensités des flux de NO<sub>x</sub> correspond au premier pic (à 5 heures) observé sur le profil moyen du cadastre optimisé (Figure 6.14). L'intensité très élevée (plus de  $4,75 \cdot 10^{12} \text{moléc.cm}^{-2} \cdot \text{s}^{-1}$ ) atteinte par le pic des émissions optimisées constitue une modification peu réaliste du cadastre *a priori*. Cependant, la modification du profil temporel avec un véritable pic entre 4 et 5 heures pour les émissions optimisées plutôt qu'un plateau pour les émissions *a priori* est semblable à la correction obtenue dans les mailles de la ville. On peut donc faire l'hypothèse que dans les mailles péri-urbaines concernées, le pic d'émissions optimisées correspond à un pic de trafic matinal mesuré mais non simulé, dû au trafic journalier pour Paris au mois d'août ou à des départs en vacances, dont l'intensité est exagérée, peut-être à cause d'un manque de représentativité des mesures.

#### 6.4.3.2 Impact sur les concentrations d'ozone

L'impact des émissions de NO<sub>x</sub> optimisées sur les concentrations d'ozone est dû à la fois à la modification de l'intensité des émissions et à celle du rapport COV sur NO<sub>x</sub>. L'inversion des émissions de NO<sub>x</sub> amène donc à la question du **régime chimique**. Or, en Île-de-France, le régime chimique dans le panache de la ville est le plus souvent proche de la transition entre NO<sub>x</sub>-limité et COV-limité, des régimes différents pouvant être établis au même moment en différents lieux (Sillman et al. 2003). La complexité de la situation fait donc des concentrations d'ozone optimisées un point de validation nécessaire des émissions optimisées.

L'impact des émissions de NO<sub>x</sub> optimisées le matin sur les concentrations d'ozone dans le panache l'après-midi est faible, comme indiqué à 15 heures sur la Figure 6.18. Les concentrations simulées lors du pic sont augmentées de 5 ppb au maximum, ce qui correspond à une modification de seulement 4,5%. Cette correction correspond à une amélioration des concentrations simulées, les concentrations *a priori* simulées à cette heure étant sous-estimées de 8% par rapport aux concentrations mesurées (voir page 99).

On peut cependant se demander si des concentrations d'ozone optimisées en aussi bon accord avec les concentrations mesurées n'auraient pas pu être obtenues sans modifier le rapport COV sur NO<sub>x</sub> à l'émission. En conservant ce rapport, les régimes chimiques établis dans le domaine avec les émissions *a priori* ne sont pas modifiés par les émissions optimisées. Les concentrations d'ozone simulées ne sont donc améliorées que si leur sous-estimation est due à une mauvaise estimation de l'intensité «absolue» des flux de COV et de NO<sub>x</sub>.

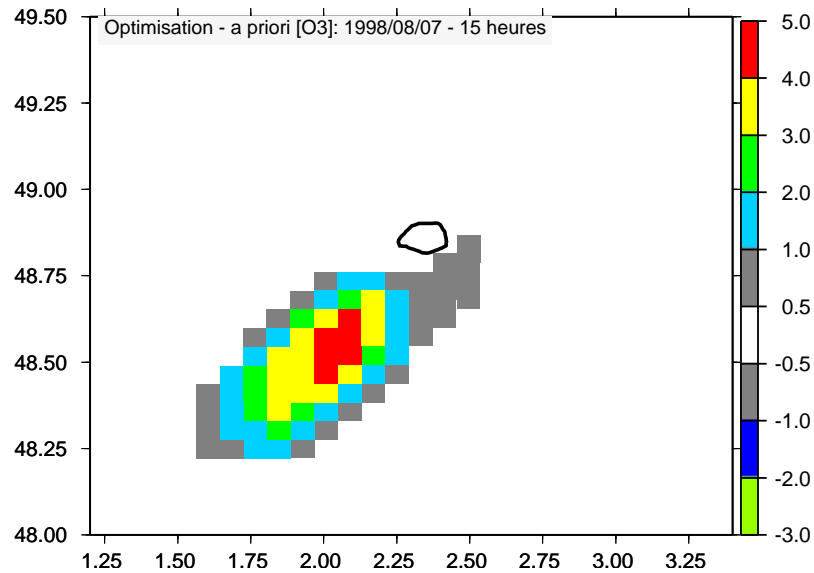


FIG. 6.18 – Différence (ppb) entre les concentrations d’ozone optimisées et a priori le 7 août 1998 à 15 heures TU. En été, TU = heure légale - 2 heures.

**Rapport COV sur  $\text{NO}_x$  conservé** Une inversion identique à celle qui a été discutée jusqu’ici mais conservant le rapport COV sur  $\text{NO}_x$  a priori a été effectuée : le même coefficient correctif a été appliqué à toutes les espèces des deux familles. Les contraintes utilisées étant toujours les concentrations de monoxyde d’azote, sur lesquelles les émissions de COV n’ont quasiment pas d’influence, les résultats obtenus pour les émissions de  $\text{NO}_x$  sont les mêmes que lors de la première optimisation à moins de 3,2%. Le profil temporel moyen des émissions totales de COV optimisées (Figure 6.19) suit donc celui des flux de  $\text{NO}_x$  de façon à maintenir le rapport COV sur  $\text{NO}_x$  à l’émission égal au rapport a priori. Les deux profils temporels présentent les mêmes caractéristiques avec un pic matinal intensifié entre 4 et 5 heures et réduit les heures suivantes. La plus grande différence en valeur absolue est obtenue à 6 heures et correspond à une diminution de 14% par rapport au flux de  $\text{NO}_x$  a priori.

L’effet de l’inversion des émissions de  $\text{NO}_x$  et de COV matinales sur les concentrations d’ozone dans le panache l’après-midi est opposé à celui de l’inversion des émissions de  $\text{NO}_x$  seules : les concentrations d’ozone simulées lors du pic à 15 heures sont diminuées de 1,5% (Figure 6.20). Bien que l’impact en soit faible, l’inversion conservant le rapport COV sur  $\text{NO}_x$  à l’émission aggrave donc la sous-estimation des concentrations d’ozone simulées. Ceci ne signifie pas que les émissions de COV a priori sont plus «justes» que les émissions obtenues ici mais que le rapport COV sur  $\text{NO}_x$  a priori ne peut pas être considéré comme une donnée à conserver.

Finalement, puisqu’on ne dispose pas de contraintes permettant d’inverser les émissions de COV indépendamment, l’optimisation des émissions de  $\text{NO}_x$  permet d’optimiser à la fois l’intensité de celles-ci et le rapport COV sur  $\text{NO}_x$ .

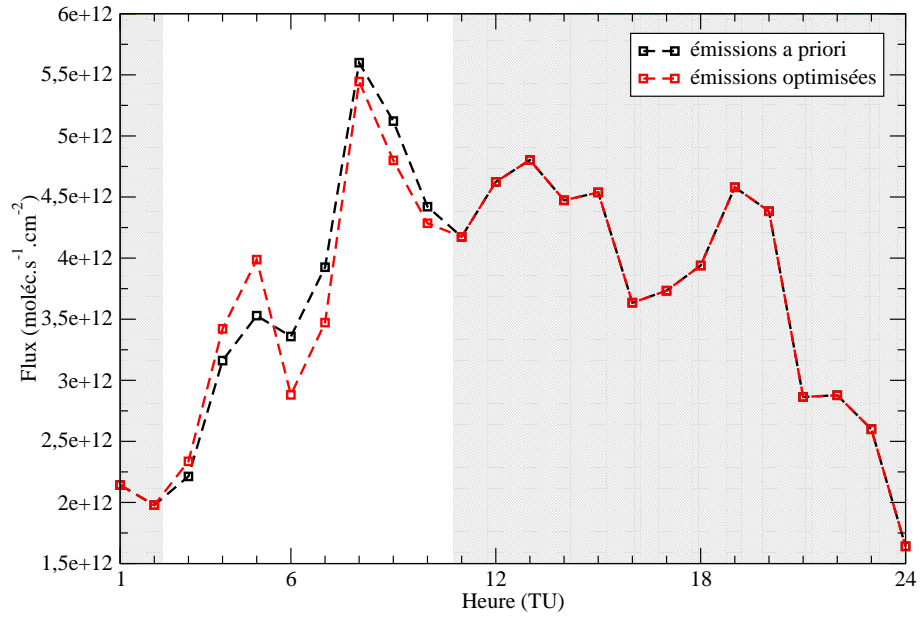


FIG. 6.19 – Profil temporel moyen sur tout le domaine du flux total d'émission de COV le 7 août 1998. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.

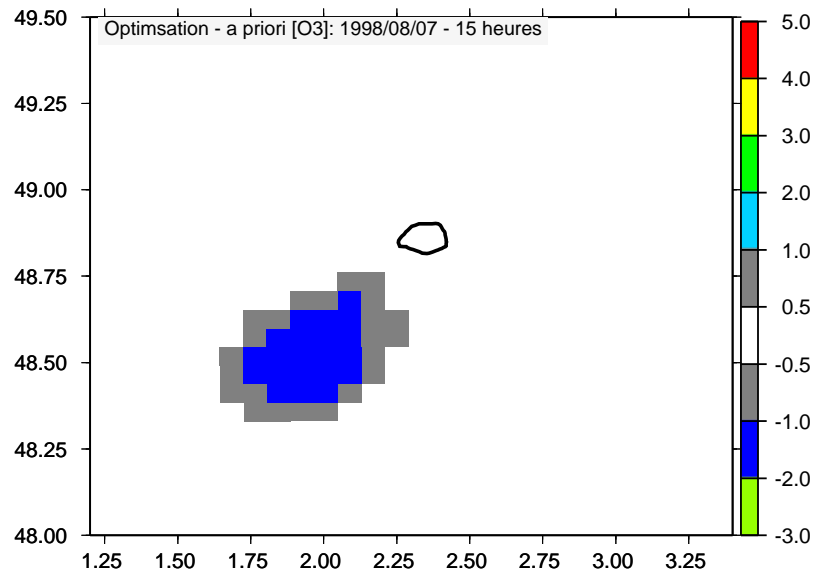


FIG. 6.20 – Différence (ppb) entre les concentrations d'ozone optimisées et a priori le 7 août 1998 à 15 heures TU. En été, TU = heure légale - 2 heures.

#### 6.4.4 Conclusion

La journée du 7 août 1998, au cours de laquelle s'est développé un épisode local de pollution photo-oxydante bien documenté par la campagne ESQUIF, a été choisie pour la première étude complète d'inversion sur un cas réel. Les études menées dans le cadre de la campagne ont indiqué une sous-estimation des concentrations d'ozone simulées dans le panache l'après-midi, probablement liée à la surestimation des émissions de  $\text{NO}_x$  *a priori* le matin en ville ou à une sous-estimation du mélange vertical.

Une rapide étude de sensibilité a permis de constater que la paramétrisation du coefficient de mélange turbulent ne pouvait expliquer les écarts entre mesures et modèle. La correction de la hauteur de couche-limite, très sous-estimée par les données de l'ECMWF par rapport aux mesures effectuées lors d'ESQUIF, a par contre permis d'expliquer un peu moins de 40% de l'écart entre concentrations de NO mesurées et simulées le matin en ville.

Une fois la hauteur de couche-limite débiaisée, les émissions de  $\text{NO}_x$  ont été optimisées. Le profil temporel des émissions optimisées présente un pic de durée inférieure au plateau du profil *a priori*. Pour les flux les plus intenses correspondant à la ville notamment, le pic matinal est réduit en durée de cinq heures, de 5 à 9 heures, à deux heures, de 8 à 9 heures. Le cadastre *a priori* ayant été élaboré sur la base du mois de juillet 1998, cette réduction de la durée du pic d'émission de  $\text{NO}_x$  matinal pour le mois d'août semble réaliste. De plus, le cadastre élaboré par AIRPARIF pour l'année 2000 à partir de données de la DREIF indique que la durée du pic du soir pour un jour ouvrable moyen est plus grande que celle du pic du matin (quatre heures contre deux, de 8 à 9 heures) (AIRPARIF 2004). Certaines caractéristiques du cadastre optimisé sont moins réalistes : un pic d'émission très intense entre 4 et 5 heures dans quatre mailles péri-urbaines paraît bien placé dans le temps mais d'une intensité surestimée.

Les émissions de  $\text{NO}_x$  optimisées le matin provoquent une augmentation de moins de 5% des concentrations d'ozone simulées dans le panache à 15 heures, les concentrations *a priori* à la même heure étant sous-estimées de 8% par rapport aux mesures. Cet impact positif est dû à la fois au changement d'intensité des émissions de  $\text{NO}_x$  et à la modification du rapport COV sur  $\text{NO}_x$  à l'émission induit par l'optimisation des flux de  $\text{NO}_x$ . La comparaison avec une optimisation de l'intensité des flux de  $\text{NO}_x$  et de COV, obtenue en répercutant les modifications des émissions de  $\text{NO}_x$  sur la famille entière des COV, a montré que les concentrations d'ozone obtenues en conservant le rapport COV sur  $\text{NO}_x$  *a priori* étaient moins proches des mesures d'une part que les concentrations optimisées précédentes et d'autre part que les concentrations *a priori*. L'optimisation de l'intensité des émissions de  $\text{NO}_x$  et du rapport COV sur  $\text{NO}_x$  à l'émission est donc plus satisfaisante étant donnée les connaissances actuelles sur les régimes chimiques en Île-de-France.



## 6.5 Du 16 au 18 juillet 1999

### 6.5.1 Choix de cette période

Les vendredi 16, samedi 17 et dimanche 18 juillet 1999 constituent la POI 6 de la campagne ESQUIF. Ces trois jours présentent les mêmes caractéristiques météorologiques que la journée du 7 août 1998 : une situation anticyclonique amenant un épisode de pollution locale. De plus, les trois jours de cette POI correspondent aux trois types de jours définis pour les émissions dues au trafic : semaine pour le vendredi, samedi et dimanche ou jour férié.

Des panaches d'ozone très concentrés (plus de 120 ppb le 17) se sont développés à des endroits différents pendant les trois journées de la POI 6.

Le 16 juillet, le panache se situe au Sud-Est de Paris et les concentrations n'y dépassent pas 80 ppb pour des concentrations au vent d'environ 50 ppb (Figure 6.21 (a)). Mesures et modèle sont en bon accord pour les valeurs des concentrations malgré un léger décalage du panache simulé vers le Sud (ESQUIF 2001; Hodzic 2002; Vautard et al. 2003).

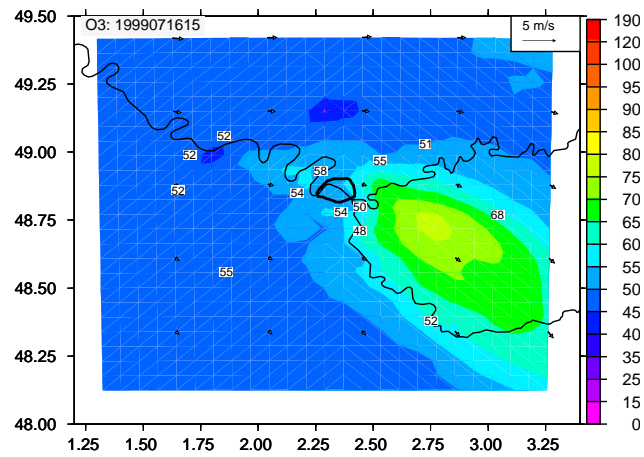
Le 17 juillet, le panache se développe à l'Ouest de Paris et les concentrations y dépassent 100 ppb pour des concentrations en amont supérieures à 60 ppb (Figure 6.21 (b)). Pour cette journée, les mesures et le modèle sont en bon accord tant pour les valeurs des concentrations que pour la position du panache (ESQUIF 2001; Hodzic 2002).

Enfin, le 18 juillet, le panache est situé près de la ville, au Nord-Ouest. Les concentrations d'ozone y sont supérieures à 100 ppb pour des concentrations au vent supérieures à 65 ppb (Figure 6.21 (c)). Les valeurs des concentrations mesurées et simulées sont en bon accord bien que le modèle situe le panache légèrement plus à l'Ouest que les mesures (ESQUIF 2001; Hodzic 2002).

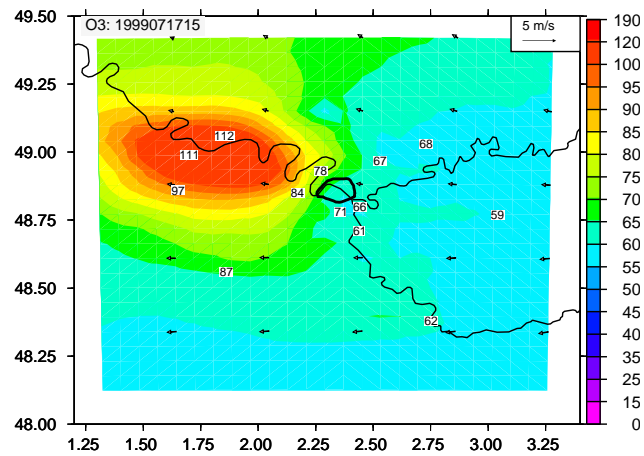
Une comparaison plus détaillée des concentrations de monoxyde d'azote mesurées et simulées indique cependant que les concentrations de NO simulées le matin en zone urbaine (Paris) ou péri-urbaine (Garches) sont très supérieures aux concentrations mesurées le samedi et le dimanche (Figure 6.22). Le pic matinal de NO simulé le samedi est plus de deux fois supérieur au pic mesuré et le pic simulé le dimanche matin n'est pas observé. La surestimation par le modèle des concentrations de NO<sub>x</sub> explique la sous-estimation des concentrations d'ozone simulées en ville (écart par exemple de 10 ppb pour le pic du samedi à Paris Sud-Ouest) et dans le panache (écart de plus de 10 ppb pour le pic du vendredi) (Figure 6.23).

Les différences entre concentrations de NO<sub>x</sub> mesurées et simulées sont particulièrement grandes le samedi et le dimanche. Or, les variations diurnes des émissions utilisées pour ces deux types de jour sont très incertaines (ESQUIF 2001). De plus, les faibles concentrations d'ozone observées en amont de la ville sont bien reproduites (ESQUIF 2001). On peut donc supposer que le rapport COV sur NO<sub>x</sub> à l'émission utilisé par le modèle joue un rôle prépondérant dans les écarts entre concentrations mesurées et simulées, en plus d'éventuelles erreurs sur la hauteur de couche-limite ou la chimie.

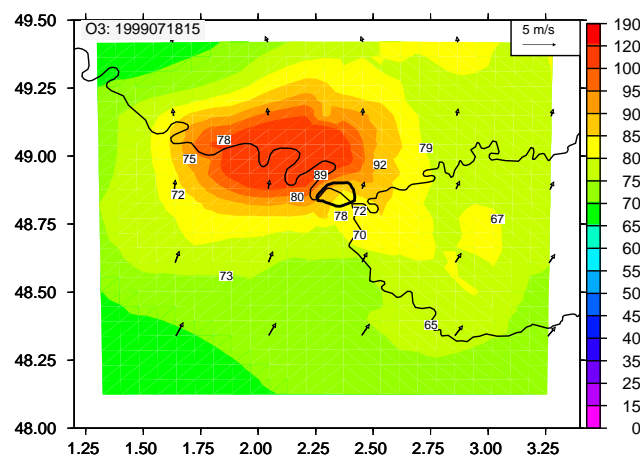
Après avoir corrigé la hauteur de couche-limite fournie au modèle à l'aide des mesures de la campagne ESQUIF, on inversera les émissions de NO<sub>x</sub> et on s'intéressera particulièrement aux caractéristiques des trois types de jours.



(a) vendredi 16



(b) samedi 17



(c) dimanche 18

FIG. 6.21 – Concentrations d' $O_3$  (en ppb) en surface les 16, 17 et 18 juillet 1999 à 15 heures TU. Les flèches noires indiquent le sens et l'intensité du vent, les étiquettes correspondent aux valeurs relevées par des stations de mesure AIRPARIF. En été, TU = heure légale - 2 heures.

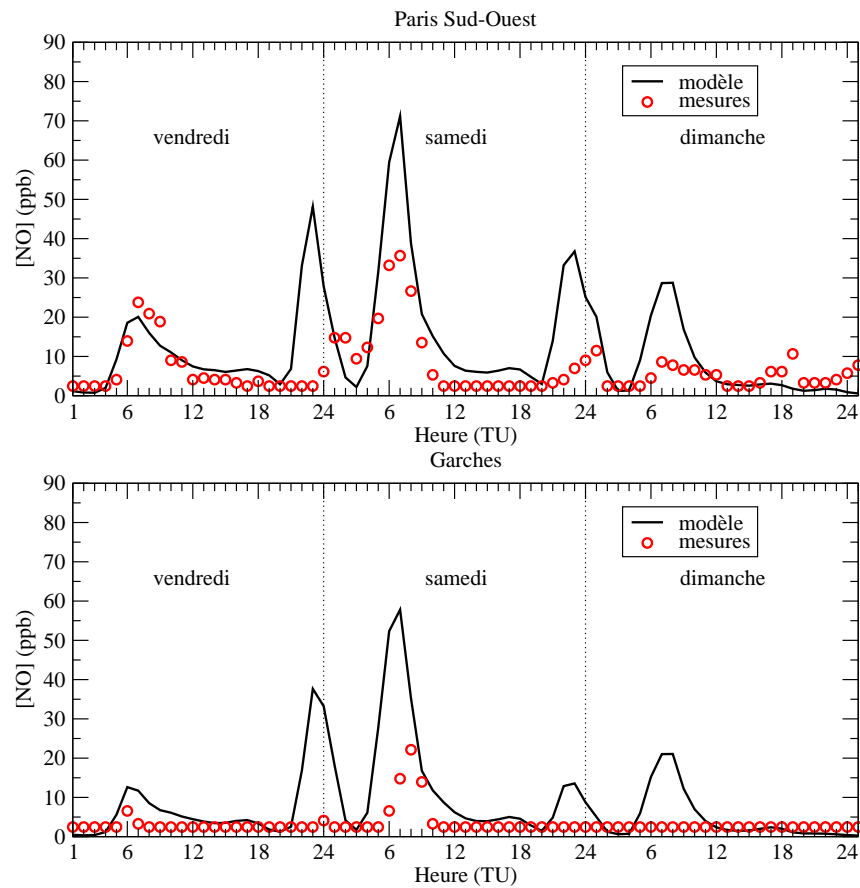


FIG. 6.22 – Comparaison entre les concentrations de monoxyde d’azote mesurées par des stations AIRPARIF et calculées par le modèle du 16 au 18 juillet 1999. En été, TU = heure légale - 2 heures.

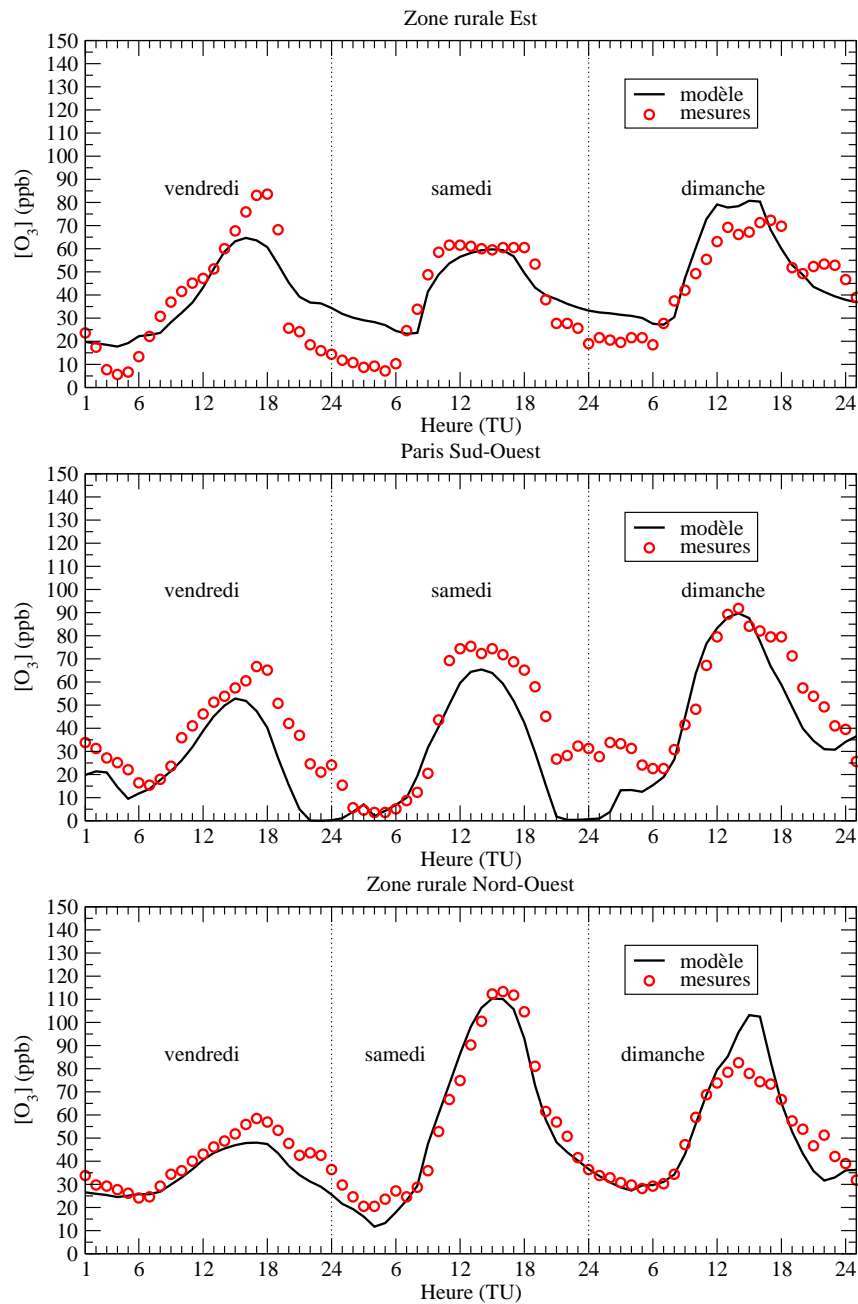


FIG. 6.23 – Comparaison entre les concentrations d’ozone mesurées par des stations AIRPARIF et calculées par le modèle du 16 au 18 juillet 1999. La Zone Rurale Nord-Ouest correspond à la station de Frémenville et la Zone Rurale Est, à la station de Saints - voir Figure 6.2 page 88 pour l’emplacement des stations. En été, TU = heure légale - 2 heures.

### 6.5.2 Inversion des émissions de $\text{NO}_x$

Pendant la POI 6, des mesures de hauteur de couche-limite ont été effectuées par le lidar de Palaiseau, en plus des sondages de Trappes. La comparaison avec les valeurs de l'ECMWF (Figure 6.24 tirée de Fochesatto et al. (2001)) indique une sous-estimation de la HCL utilisée par le modèle.

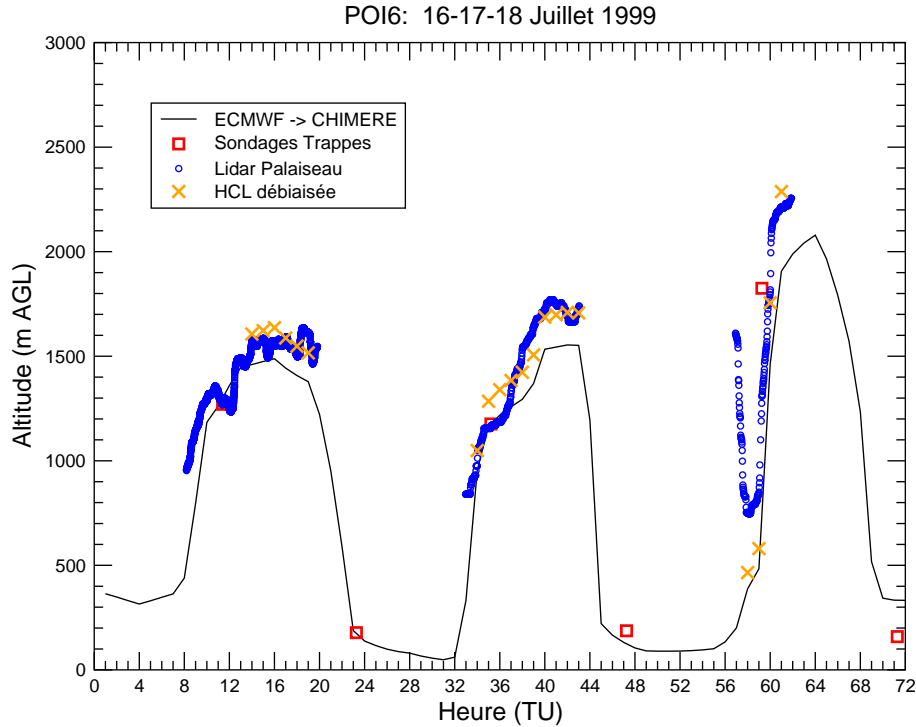


FIG. 6.24 – Hauteur de couche-limite pendant la POI 6 d'ESQUIF : comparaison entre les mesures et les valeurs fournies au modèle par l'ECMWF. En été, TU = heure légale - 2 heures.

Avant l'inversion des émissions, la HCL a donc été corrigée d'après mesures du lidar de Palaiseau : elle a été augmentée de 10% le 16 juillet entre 14 et 19 heures, de 10% le 17 juillet entre 10 et 19 heures et de 20% le 18, entre 10 et 19 heures (Figure 6.24).

Seuls les flux de  $\text{NO}_x$  sont inversés donc le rapport COV sur  $\text{NO}_x$  à l'émission est modifié. L'agrégation spatiale est effectuée avec une tolérance  $\tau_E \geq 25\%$  qui permet de regrouper, pour la plupart des heures inversées, les quatre flux correspondant à la ville au sein d'une même zone. Deux cycles krigeage-optimisation sont enchaînés.

**Moyenne sur tout le domaine** La tendance globale donnée par l'optimisation aux flux d'émission de  $\text{NO}_x$  est indiquée par le profil temporel moyen sur tout le domaine (Figure 6.25). Par rapport au cadastre *a priori*, l'intensité des pics matinaux du cadastre optimisé (de 4 à 11 heures) est peu modifiée le vendredi et réduite pour les deux types de jours «samedi» et «dimanche ou jour férié». Les modifications sont inférieures à 2% (en valeur absolue) le vendredi ; la réduction atteint -23% à 6 heures le samedi et -28,6% à 6 heures le dimanche. La masse moyenne émise au cours de la journée est inchangée le

vendredi et réduite de 3,2% le samedi et 4,7% le dimanche.

Ces corrections conservent le profil temporel moyen des émissions le vendredi ; la durée du pic du samedi est réduite de cinq heures (de 5 à 9 heures) *a priori* à trois heures (de 7 à 9 heures). Enfin, le dimanche, le profil temporel moyen *a priori*, comportant un pic matinal culminant à 7 heures, est modifié par un déficit par rapport au cadastre *a priori* à 8 heures, qui devra être examiné plus particulièrement.

Les réductions de l'intensité et les modifications de l'allure du profil temporel des flux d'émission sont beaucoup plus importantes pour le samedi et le dimanche que pour le vendredi.

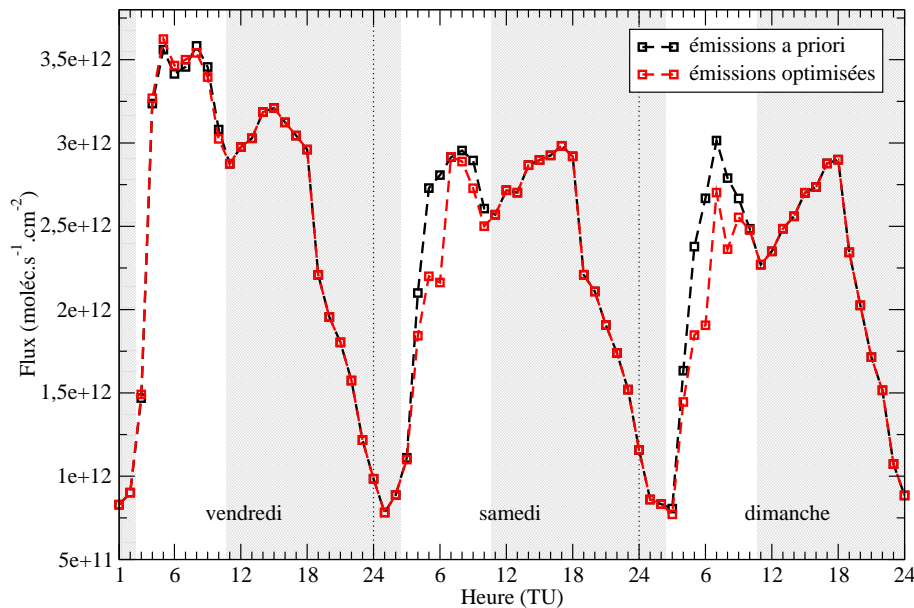


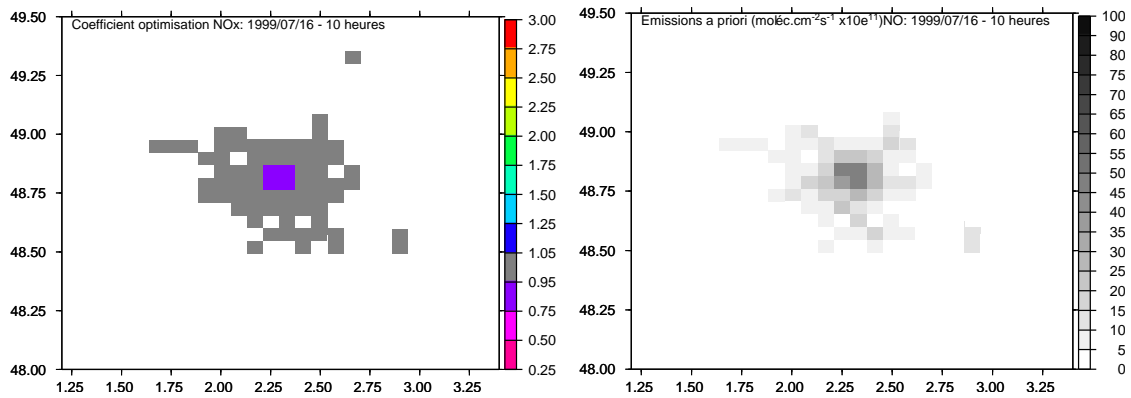
FIG. 6.25 – Profil temporel moyen sur tout le domaine du flux d'émission de NO du 16 au 18 juillet 1999. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.

**Répartition spatiale** Afin de déterminer si les réductions moyennes concernent tous les flux, la répartition spatiale des corrections est examinée à chaque heure inversée. Les heures qui présentent les plus grandes disparités dans les coefficients de correction sont représentées sur la Figure 6.26.

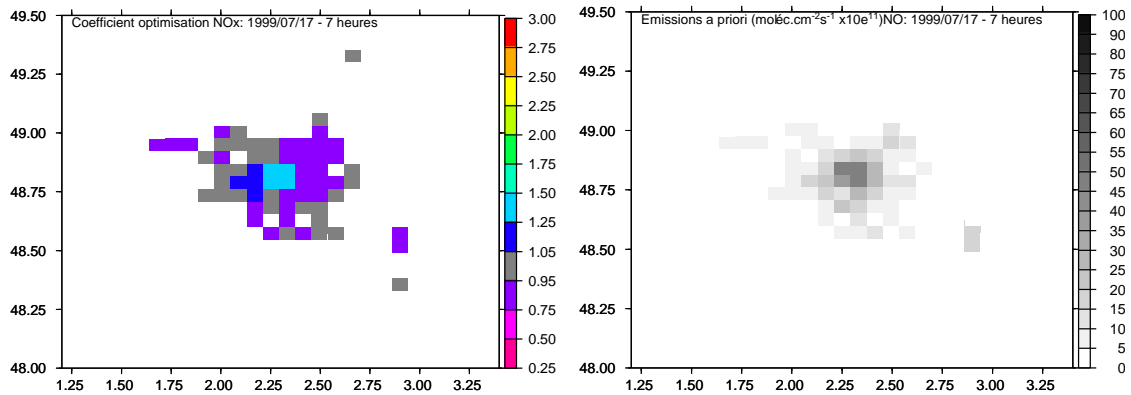
**Le vendredi** à 10 heures, les flux les plus intenses correspondant à la ville sont multipliés par 0,75 à 0,95. Les flux extérieurs à Paris, moins intenses, sont peu modifiés, le coefficient de l'optimisation restant compris entre 0,95 et 1,05. Ainsi la stabilité moyenne de l'intensité des émissions du vendredi matin correspond-elle à de faibles modifications (inférieures à 5%) de la majorité des flux.

**Le samedi** à 7 heures, les flux les plus intenses correspondant à la ville sont multipliés par 1,25 à 1,5 alors que les flux intenses extérieurs à Paris situés en aval sont multipliés par 1,05 à 1,25. Le coefficient de correction des autres flux inversés reste compris entre 0,75 et 1,05. Le pic du profil temporel moyen correspond donc à l'augmentation, dans le cadastre optimisé, de certains flux très intenses alors que la majorité des flux d'émissions

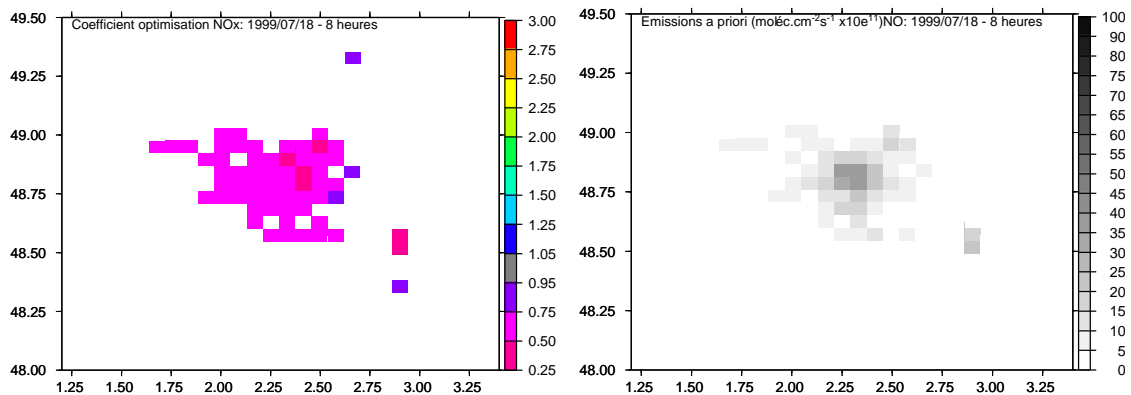
sont diminués par rapport au cadastre *a priori*.



(a) vendredi 16 à 10 heures TU



(b) samedi 17 à 7 heures TU



(c) dimanche 18 à 8 heures TU

FIG. 6.26 – Répartition spatiale des corrections pour les  $NO_x$  et intensité des émissions de NO (représentant 90% des flux de  $NO_x$ ) les 16, 17 et 18 juillet 1999. En été, TU = heure légale - 2 heures.

**Le dimanche** à 8 heures, les corrections comprises entre 0,5 et 0,75 sont réparties de façon très homogène sur le cadastre. Le déficit apparaissant sur le profil temporel moyen correspond donc à une réduction de -25 à -50% de tous les flux inversés, certains étant réduits de plus de 50% (coefficients compris entre 0,25 et 0,5).

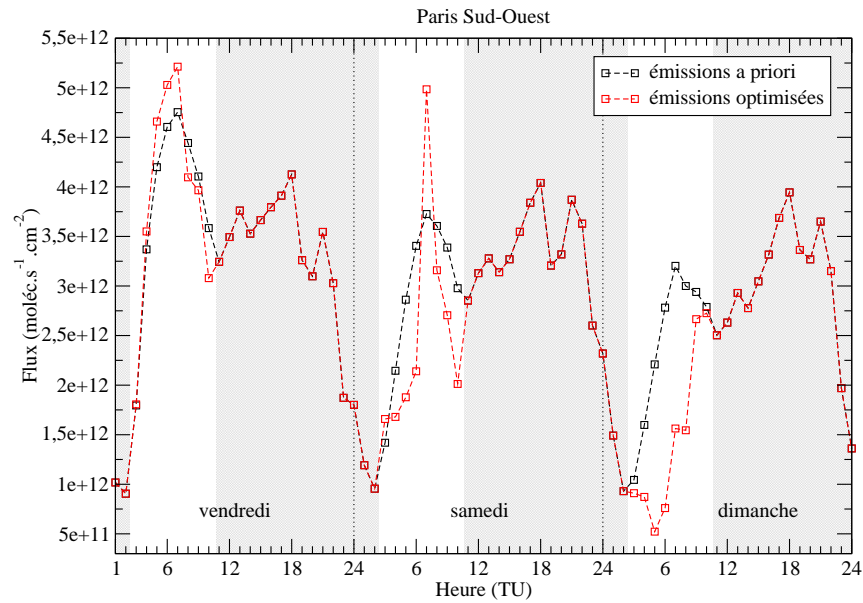
**Différents lieux** Pour les journées du samedi et du dimanche, le profil temporel moyen des émissions recouvre donc des situations différentes selon les lieux : dans Paris et aux endroits où les flux sont moins intenses. L'évolution temporelle des flux d'émissions et des concentrations de monoxyde d'azote dans la ville est représentée sur la Figure 6.27 pour la maille correspondant à Paris Sud-Ouest (représentative de la ville) et sur la Figure 6.28 pour la maille de Tremblay.

Comme indiqué précédemment, le profil temporel du **vendredi** est conservé partout, les flux les plus intenses correspondant à la ville ayant été augmentés de façon plus importante pour seulement trois heures (entre 5 et 8 heures). Le profil temporel des concentrations de NO est lui aussi inchangé, seule l'intensité est modifiée : en ville, le pic optimisé atteint 20 ppb contre 22 ppb simulés *a priori* pour une mesure de 23,7 ppb et une contrainte de 21,5 ppb. On peut supposer que la réduction du pic de NO<sub>x</sub> le matin du vendredi 16 juillet correspond à une réduction du trafic due à des départs en «long week-end» du mercredi 14 (férié) au dimanche 17.

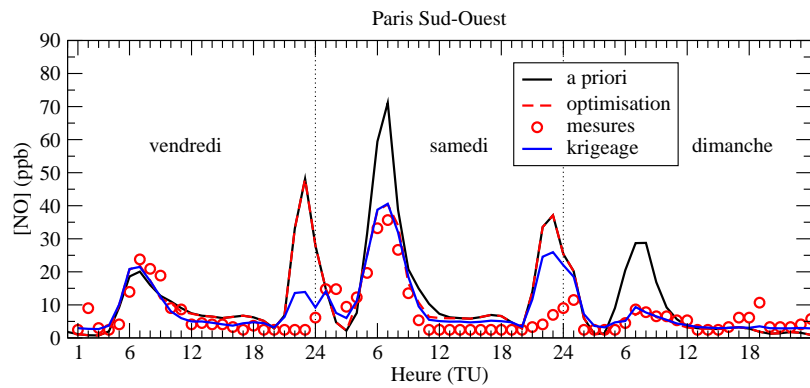
Les deux profils temporels font apparaître la grande différence constatée **le samedi** : les émissions optimisées dans la maille urbaine constituent un pic plus intense mais plus court que dans le cadastre *a priori* (5 au lieu de  $3,7 \cdot 10^{12}$  moléc.cm<sup>-2</sup>.s<sup>-1</sup> à 7 heures) alors que dans la maille péri-urbaine, le pic matinal a disparu du cadastre optimisé (moins de 1,5 au lieu de plus de  $1,7 \cdot 10^{12}$  moléc.cm<sup>-2</sup>.s<sup>-1</sup> de 5 à 7 heures). Les concentrations de NO optimisées suivent le même profil temporel que les concentrations simulées *a priori*. Elles sont cependant très inférieures à ces dernières : en ville, la concentration du pic matinal est diminuée de plus de 40% (71 ppb *a priori* et 40 ppb après optimisation) et dans la maille péri-urbaine, elle est divisée par deux (de 39 ppb *a priori* à 20 ppb après optimisation). Les corrections apportées par l'optimisation aux émissions du samedi sont donc importantes ; à ce stade cependant, elles ne comportent aucun élément irréaliste. On remarquera que la différence entre le vendredi et le samedi est réduite par l'optimisation, ce qui pourrait correspondre à une différence de trafic moins marquée entre jour de semaine et jour ouvrable de week-end en période de vacances estivales.

Le profil temporel du **dimanche** en ville indique que le pic de concentration de NO qui était simulé avec le cadastre *a priori* entre 7 et 8 heures mais non observé disparaît avec le cadastre optimisé. Le pic d'émissions correspondant, qui dans le cadastre *a priori* dépassait  $2,7 \cdot 10^{12}$  moléc.cm<sup>-2</sup>.s<sup>-1</sup> de 6 à 10 heures, n'existe plus dans le cadastre optimisé, les flux atteignant juste  $2,7 \cdot 10^{12}$  moléc.cm<sup>-2</sup>.s<sup>-1</sup> à 9 et 10 heures. L'évolution temporelle des émissions dans la maille de Tremblay est la même que dans Paris, les émissions optimisées entre 5 et 9 heures étant très inférieures aux émissions *a priori* (de -14 à -73%) sauf à 7 heures. À cette heure seulement, les flux optimisés ne sont inférieurs que de 12,6% aux flux réels et le profil temporel optimisé comporte donc un pic d'émission d'une heure. De même, les concentrations de NO optimisées sont très faibles par rapport aux concentrations *a priori* : le pic atteint 13,6 ppb à 8 heures au lieu de plus de 18 ppb entre 7 et 8 heures. Les corrections obtenues pour les flux de la ville et les flux péri-urbains permettent de remédier à un défaut de réalisme du cadastre *a priori* en éliminant un pic non





(a) Émissions a priori et optimisées

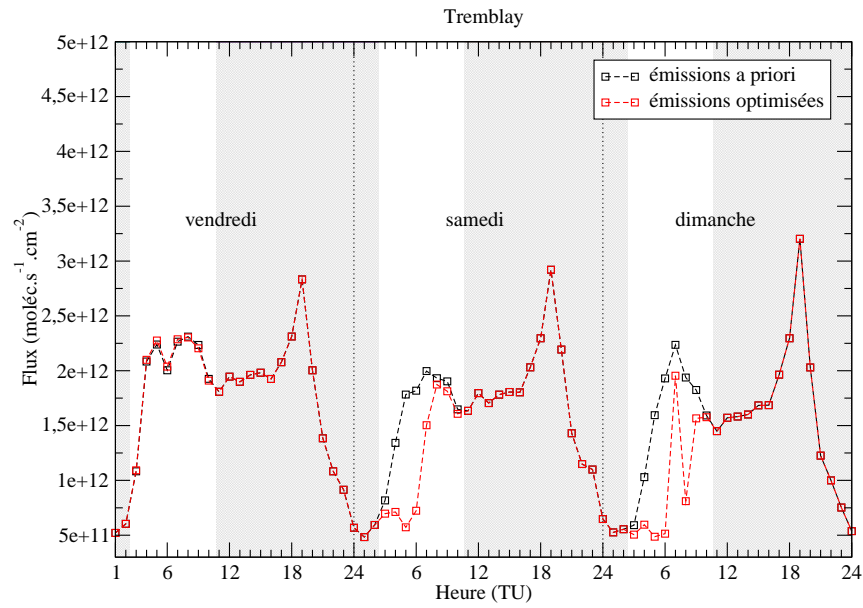


(b) Concentrations de monoxyde d'azote a priori, optimisées et mesurées

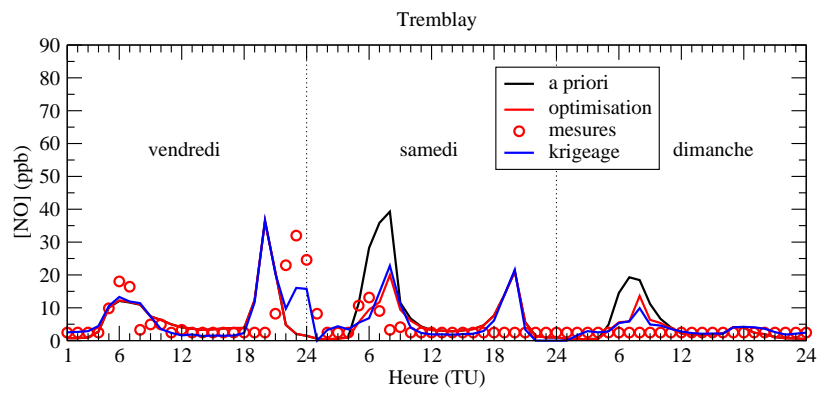
FIG. 6.27 – Évolutions temporelles des émissions et des concentrations de monoxyde d'azote du 16 au 18 juillet 1999 dans la maille Paris Sud-Ouest. Les concentrations d'une heure donnée résultent des émissions de l'heure précédente. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.

observé et ne comportent en elles-mêmes aucun élément irréaliste. Pour les mailles péri-urbaines cependant, le pic d'émission du profil temporel optimisé (à 7 heures) constitue une oscillation peu réaliste.

Finalement, pour le samedi et le dimanche, les deux types de jours comportant une grande incertitude sur le profil d'émission utilisé pour élaborer le cadastre *a priori*, les corrections apportées par l'optimisation consistent en d'importantes augmentations (le samedi en ville) ou réductions (le dimanche en ville, le samedi en périphérie de l'agglomération)



(a) Émissions a priori et optimisées



(b) Concentrations de monoxyde d'azote a priori, optimisées et mesurées

FIG. 6.28 – Évolutions temporelles des émissions et des concentrations de monoxyde d'azote du 16 au 18 juillet 1999 dans la maille de Tremblay. Les concentrations d'une heure donnée résultent des émissions de l'heure précédente. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.

de l'intensité des flux émis. La plupart de ces modifications paraissent réalistes (on peut notamment en proposer des explications en terme de trafic). Les émissions optimisées du dimanche présentent cependant des caractéristiques irréalistes. Ceci peut être lié aux écarts très importants entre concentrations de monoxyde d'azote mesurées et simulées constatés ce jour-là tant en ville que dans certaines mailles péri-urbaines. La tentative de réduction de ces écarts lors de l'optimisation a pu conduire à l'apparition d'oscillations dans les profils temporels des flux d'émission des mailles péri-urbaines.

### 6.5.3 Conclusion

Les vendredi 16, samedi 17 et dimanche 18 juillet 1999, au cours desquels s'est développé un épisode de pollution photo-oxydante bien documenté par la campagne ESQUIF, ont été choisis pour une première étude comparative entre les trois types de jours utilisés pour l'élaboration du cadastre (semaine, samedi et dimanche ou jour férié). Les études menées dans le cadre de la campagne ont indiqué une forte surestimation des concentrations de monoxyde d'azote simulées le matin en ville et dans les zones péri-urbaines, particulièrement marquée le samedi et le dimanche. Étant donnée la forte incertitude liée aux profils temporels diurnes de ces deux types de jours, on a pu supposer que le rapport COV sur  $\text{NO}_x$  à l'émission était un paramètre à optimiser.

Une fois la hauteur de couche-limite débiaisée, une inversion portant sur les seules émissions de  $\text{NO}_x$  a été effectuée. Les corrections obtenues pour les trois types de jours sont très différentes. L'allure du profil temporel du vendredi est conservée, l'intensité des émissions étant modifiée de moins de 5% pour la majorité des flux inversés à 25% pour quelques flux plus intenses. Le samedi, les résultats obtenus diffèrent selon les lieux : les émissions optimisées forment en ville un pic plus intense d'environ 35% que le pic *a priori* mais d'une durée plus courte (deux heures au lieu de cinq) ; à l'extérieur de Paris, les flux d'émission sont réduits de telle sorte que le pic matinal *a priori* disparaît complètement du cadastre optimisé. Les corrections obtenues sont importantes mais ne comportent aucun élément irréaliste à ce stade. Enfin, le dimanche, le pic de concentration de NO qui était simulé en ville avec le cadastre *a priori* mais non observé disparaît avec le cadastre optimisé. Cependant, les corrections obtenues pour les mailles péri-urbaines, qui amènent un pic d'émission isolé à 7 heures, paraissent peu réalistes. Ce manque de réalisme peut être dû à la réduction d'un très grand écart entre concentrations mesurées et simulées (7,5 fois plus élevées à 8 heures lors du pic simulé mais non observé).

La comparaison des deux vendredis, le 7 août 1998 et le 16 juillet 1999, indique des tendances différentes : dans le premier cas, le profil temporel des émissions est modifié (durée du pic réduite) alors que dans le deuxième cas, il est conservé. Cependant, dans tous les cas, le réalisme des corrections obtenues par l'optimisation est plus satisfaisant pour les flux intenses, en ville, c'est-à-dire là où sont concentrées de nombreuses mesures, que pour les flux périphériques, moins intenses et situés dans des zones où le réseau de mesure est moins dense. On retrouve donc les limites de la méthodologie malgré l'utilisation du krigeage qui permet de répartir au mieux l'information disponible dans les concentrations mesurées.

Les deux études réalisées sur des cas de pollution photo-oxydante bien documentés ont permis de montrer que la méthodologie d'inversion était applicable à des cas réels. Les hypothèses posées lors de la validation sur des cas académiques sont donc raisonnablement réalistes. Cependant, les résultats obtenus pour ces deux études ne sont pas généralisables à d'autres périodes. En effet, les cadastres optimisés comportent des caractéristiques particulières à chacun des cas, ce qui ne permet pas de dégager des tendances valables pour une semaine ou un mois entiers. Il s'avère donc nécessaire de mener une étude d'inversion à plus long terme.

## 6.6 Inversions climatologiques

### 6.6.1 Choix de la période

Cette étude climatologique s'étend sur les mois de juillet et août 1998 et 1999. Cette période de deux fois deux mois permet d'englober les POI 2 et 6 d'ESQUIF, étudiées individuellement dans les Sections précédentes.

Grâce à cette étude sur deux étés, il est possible déterminer si l'optimisation des émissions permet d'améliorer systématiquement les concentrations simulées. Parmi tous les paramètres susceptibles d'influencer l'inversion des flux d'émission, on examinera particulièrement les conditions aux limites et la hauteur de couche-limite, qui ont, potentiellement, le plus fort impact (Menut 2003; Menut et al. 2000). Enfin, on pourra étudier la variabilité temporelle des émissions selon les différents jours, notamment les éventuelles différences entre les jours de la semaine, non représentées dans le cadastre *a priori*.

### 6.6.2 Inversion des émissions de $\text{NO}_x$

On n'utilise aucune information autre que les mesures du réseau AIRPARIF pour l'inversion des émissions de  $\text{NO}_x$ . La HCL n'est donc pas débiaisée avant l'inversion. Cependant, l'utilisation, pour des prévisions sur de longues périodes ([www.prevair.org/fr/](http://www.prevair.org/fr/)), des valeurs de hauteur de couche-limite fournies au modèle par l'ECMWF montre qu'en moyenne, sur l'ensemble du domaine, la variabilité de la HCL est bien reproduite. L'influence, sur les flux optimisés, d'éventuelles corrections de la HCL grâce à des mesures sera examinée pour les jours des POI 2 et 6 d'ESQUIF étudiés précédemment. L'agrégation spatiale est effectuée avec une tolérance  $\tau_E \geq 25\%$ , qui permet de regrouper, pour la plupart des heures inversées, les quatre flux correspondant à la ville au sein d'une même zone. Deux cycles krigeage-optimisation sont enchaînés, de façon à améliorer la qualité des résultats par rapport à un seul cycle mais sans utiliser un temps de calcul trop long (voir page 66).

#### Impact sur les concentrations

Le biais, la corrélation et la moyenne quadratique sur les mois de juillet et août sont calculés pour les concentrations *a priori* et optimisées de monoxyde d'azote et d'ozone. Pour quantifier l'impact de l'optimisation sur les concentrations, les écarts entre les valeurs obtenues avec le cadastre *a priori* et les valeurs obtenues avec le cadastre optimisé sont indiquées en pourcentage dans les Tableaux 6.4 et 6.5 : par exemple, pour le biais, on a 
$$\Delta \text{biais} = \frac{\text{biais}_{\text{apriori}} - \text{biais}_{\text{optimisé}}}{\text{biais}_{\text{apriori}}} \times 100.$$

**Le monoxyde d'azote** Les concentrations de monoxyde d'azote analysées ont été utilisées comme contraintes pour l'inversion. Les concentrations optimisées doivent donc être beaucoup plus proches des mesures que les concentrations simulées *a priori*, tant en intensité qu'en phasage temporel. Dans les stations urbaines (partie supérieure du Tableau 6.4), ceci est confirmé par les variations du biais, centrées sur -33,2% et les variations de l'erreur moyenne, centrées sur -15%, qui indiquent que les écarts en intensité sont diminués après optimisation. Les fortes augmentations du biais à Paris Sud-Est et Sud-Ouest

(+123,9 à +845,9%, Tableau 6.4) correspondent à des valeurs initiales très faibles, inférieures à 0,8 ppb en valeur absolue : ces biais restent donc inférieurs à 6,5 ppb en valeur absolue. Les variations de la corrélation, centrées sur +28,4% dans les stations urbaines et améliorées de 1,6 à 6,3% dans les stations rurales (Tableau 6.4), confirment que le phasage temporel des mesures est mieux reproduit par le modèle après optimisation.

**L’ozone** Comme l’a montré l’étude du 7 août 1998 (voir page 109), les concentrations d’ozone peuvent servir de première validation aux résultats de l’optimisation. L’étude climatologique effectuée sur les concentrations simulées *a priori* (voir page 90) indiquait que les concentrations d’ozone étaient bien reproduites par le modèle, tant en intensité qu’en phasage temporel, malgré une légère surestimation (médiane de +2,9 ppb pour les stations urbaines). Les variations du biais, centrées sur -1,6% et de l’erreur moyenne centrées, sur -2,3%, indiquent que l’optimisation permet de réduire encore les écarts en intensité entre concentrations simulées et mesurées (Tableau 6.5). Le fort pourcentage obtenu pour le biais à Tremblay correspond au passage d’un biais quasiment nul (inférieur à 0,02 ppb) à un biais de l’ordre de -0.2 ppb. La corrélation est également améliorée (variations centrées sur +1,5%, Tableau 6.5).

Finalement, les flux d’émissions optimisés permettent une meilleure reproduction des concentrations de monoxyde d’azote et d’ozone simulées sur deux étés : par rapport au cadastre *a priori*, les valeurs du biais et de l’erreur moyenne sont diminuées et la corrélation est augmentée.

<i>Statistiques NO</i>	<i>été 1998</i>					
<b>Site</b>	biais	$\Delta$ biais	rms	$\Delta$ rms	cor	$\Delta$ cor
zones urbaines						
Argenteuil	-4.5	-22.3	11.7	-22.8	0.5	+49.1
Bobigny	-3.5	-40.9	14.3	<b>-15.0</b>	0.49	+35.1
Champs-sur-Marne	-7.7	-22.2	18.1	-22.1	0.37	+86.0
Créteil	-4.0	-41.2	15.2	-36.7	0.35	+126.6
Évry	x	x	x	x	x	x
Garches	x	x	x	x	x	x
Gennevilliers	-2.9	<b>-33.2</b>	11.4	-27.2	0.53	+48.
La Défense	-5.7	-19.6	13.6	-9.4	0.52	+16.7
Mantes	-4.6	-9.7	8.1	-16.2	0.41	+63.7
Melun	-7.6	-1.7	12.3	-2.1	0.21	+53.6
Montgeron	-1.5	-68.8	7.8	-24.7	0.4	+95.
Paris Nord-Est	-2.1	-65.7	12.4	+3.1	0.63	+2.3
Paris Nord-Ouest	-2.7	-42.	14.9	-7.1	0.48	+23.6
Paris Sud-Est	0.8	+164.0	13.5	-30.1	0.53	+55.9
Paris Sud-Ouest	0.8	+123.9	9.1	+7.	0.63	+12.6
Saint-Denis	-3.6	-38.6	13.7	-24.9	0.62	+27.3
Tremblay	-1.1	-44.5	10.6	-16.1	0.40	+81.2
Versailles	-2.3	-42.0	7.5	+9.5	0.43	<b>+28.4</b>
Vitry	-4.6	-27.7	20.1	-32.8	0.41	+93.3
zones rurales						
zone rurale Sud-Est	x	x	x	x	x	x
zone rurale Sud-Ouest	-1.9	-0.5	2.5	-0.1	0.2	+6.3
<i>Statistiques NO</i>	<i>été 1999</i>					
<b>Site</b>	biais	$\Delta$ biais	rms	$\Delta$ rms	cor	$\Delta$ cor
zones urbaines						
Argenteuil	-3.5	-12.2	8.6	-12.7	0.67	+12.9
Bobigny	-4.3	-28.5	13.8	-18.2	0.56	+23.5
Champs-sur-Marne	x	x	x	x	x	x
Créteil	-3.9	-20.2	12.1	-15.3	0.43	+49.2
Évry	-5.8	-14.	12.5	-10.8	0.43	+33.1
Garches	-0.1	+420.5	9.5	-13.2	0.44	+48.9
Gennevilliers	-3.2	-26.7	11.6	-27.6	0.65	+27.6
La Défense	x	x	x	x	x	x
Mantes	-2.7	-5.8	5.2	-9.7	0.52	+25.2
Melun	-2.9	-1.3	4.9	-1.4	0.44	+6.8
Montgeron	-2.7	-14.	5.5	-0.3	0.55	+3.3
Paris Nord-Est	-1.3	-84	10.2	-12.6	0.68	+12.7
Paris Nord-Ouest	-1.6	-87.5	9.7	-22.1	0.66	+26.6
Paris Sud-Est	0.2	+845.9	9.6	-12.3	0.58	+35.2
Paris Sud-Ouest	-0.7	-37.7	7.6	-4.3	0.69	+15.0
Saint-Denis	-2.2	-46.2	9.2	-19.6	0.69	+13.6
Tremblay	-1.9	-62.2	11.9	-22.3	0.39	+81.6
Versailles	-2.1	-14.4	5.9	-2.1	0.62	+5.9
Vitry	-3.2	-28.3	12.2	-24.3	0.49	+53.5
zones rurales						
zone rurale Sud-Est	-1.9	-0.2	2.1	-0.1	0.2	+1.6
zone rurale Sud-Ouest	-1.9	-0.4	2.5	-0.3	0.25	+3.3

TAB. 6.4 – Valeurs des variables statistiques (biais, moyenne quadratique (rms) et corrélation (cor)) calculées sur les mois de juillet et août en 1998 et 1999 pour les concentrations de NO (en ppb) a priori dans les différentes mailles où se trouvent des stations de mesure (x = pas de mesure à cette période) et écarts (en % de la valeur a priori) entre les valeurs calculées pour les concentrations de NO a priori et optimisées.

$$\Delta X = \frac{X_{\text{apriori}} - X_{\text{optimisé}}}{X_{\text{apriori}}} \times 100$$

<i>Statistiques ozone</i>	<i>été 1998</i>					
<b>Site</b>	biais	$\Delta$ biais	rms	$\Delta$ rms	cor	$\Delta$ cor
zones urbaines						
Champs-sur-Marne	7.9	-8.9	14.6	-3.4	0.77	+1.6
Créteil	4.6	-10.5	13.3	-2.1	0.77	+2.0
Garches	x	x	x	x	x	x
Gennevilliers	x	x	x	x	x	x
Mantes	4.3	-6.5	11.6	-3.3	0.80	+1.5
Melun	x	x	x	x	x	x
Montgeron	2.5	-22.0	12.5	-2.7	0.8	+1.9
Paris Nord-Est	1.2	-32.8	11.2	-1.4	0.81	<b>+1.5</b>
Paris Nord-Ouest	5.3	-7.0	12.2	-2.1	0.80	+1.7
Paris Sud-Est	-3.1	+9.4	11.6	-0.7	0.82	+1.3
Paris Sud-Ouest	1.4	-23.6	11.8	+0.3	0.79	+1.1
Tremblay	-1.9	+28.5	12.8	-2.4	0.78	+2.2
Vitry	0.8	+10.4	9.5	-12.1	0.81	+6.
zones rurales						
zone rurale Est	x	x	x	x	x	x
zone rurale Nord-Est	0.0	+1587.1	11.8	-1.4	0.77	+0.6
zone rurale Nord-Ouest	8.5	<b>-1.6</b>	11.3	-1.1	0.84	+0.4
zone rurale Ouest	x	x	x	x	x	x
zone rurale Sud-Est	x	x	x	x	x	x
zone rurale Sud-Ouest	-1.4	-4.9	10.4	+0.2	0.85	0.0
<i>Statistiques ozone</i>	<i>été 1999</i>					
<b>Site</b>	biais	$\Delta$ biais	rms	$\Delta$ rms	cor	$\Delta$ cor
zones urbaines						
Champs-sur-Marne	x	x	x	x	x	x
Créteil	7.8	-4.2	13.6	-2.7	0.77	+1.6
Garches	-0.7	+48.5	12.5	<b>-2.3</b>	0.78	+1.9
Gennevilliers	3.6	-9.7	11.9	-3.4	0.81	+1.6
Mantes	1.9	-9.3	11.1	-2.4	0.84	+1.0
Melun	6.1	-1.2	12.0	-0.8	0.82	+0.3
Montgeron	5.6	-4.4	12.1	-2.3	0.8	+1.1
Paris Nord-Est	0.2	+79.8	11.2	-4.7	0.8	+2.7
Paris Nord-Ouest	3.7	-12.6	11.4	-4.7	0.8	+2.6
Paris Sud-Est	-2.2	+20.7	12.4	-3.2	0.77	+2.9
Paris Sud-Ouest	-0.8	+49.9	11.6	-3.8	0.78	+2.8
Tremblay	-0.6	+48.3	12.2	-2.4	0.78	+1.6
Vitry	1.2	-33.2	12.3	-3.7	0.76	+2.5
zones rurales						
zone rurale Est	4.	-1.8	10.2	-0.2	0.83	0.0
zone rurale Nord-Est	-2.3	+4.3	10.9	-0.8	0.78	+0.7
zone rurale Nord-Ouest	-0.3	+53.	9.7	-2.6	0.85	+1.0
zone rurale Ouest	3.7	-1.4	10.0	-1.4	0.86	+0.5
zone rurale Sud-Est	4.9	+0.3	11.1	0.0	0.83	0.0
zone rurale Sud-Ouest	-2.8	-1.1	11.5	-0.3	0.84	+0.2

TAB. 6.5 – Valeurs des variables statistiques (biais, moyenne quadratique (rms) et corrélation (cor)) calculées sur les mois de juillet et août en 1998 et 1999 pour les concentrations d’ozone (en ppb) a priori dans les différentes mailles où se trouvent des stations de mesure (x = pas de mesure à cette période) et écarts (en % de la valeur a priori) entre les valeurs calculées pour les concentrations d’ozone a priori et optimisées.

$$\Delta X = \frac{X_{\text{apriori}} - X_{\text{optimisé}}}{X_{\text{apriori}}} \times 100$$

### Influence d'autres paramètres sur l'optimisation des émissions

Parmi tous les paramètres du modèle, les conditions aux limites et la hauteur de couche-limite ont potentiellement une influence importante (Menut 2003; Menut et al. 2000) et très différente selon les jours sur l'optimisation des émissions.

**Les conditions aux limites** Imposées par une simulation CHIMERE à l'échelle continentale (voir page 44), elles sont considérées comme correctes (hypothèse de travail réaliste, voir page 97). Cependant, la quantité d'information utile à l'inversion des émissions locales contenue dans les concentrations mesurées dépend de leur intensité. Cette situation est illustrée sur la Figure 6.29. Pour cette illustration, on a représenté (sur l'axe vertical de la Figure 6.29) les concentrations (en unités arbitraires) dues aux masses d'air advectées dans le domaine ( $C_{import}$ ), les concentrations réelles dues aux émissions ( $E_{vraies}$ ) et les concentrations simulées avec le cadastre *a priori* ( $E_{a priori}$ ). Les concentrations mesurées ( $C_{mesure}$ ) correspondent à la somme des concentrations dues aux conditions aux limites et des concentrations dues aux émissions locales. Les concentrations simulées ( $C_{simu}$ ) correspondent à la somme des concentrations dues aux conditions aux limites et des concentrations simulées avec les émissions *a priori*. Finalement, pour des flux d'émission locaux identiques, l'écart entre concentrations mesurées et simulées ( $\Delta_{fg}$ ) dû à des défauts du cadastre *a priori* représente une proportion des concentrations simulées d'autant plus faible que les concentrations importées dans le domaine sont élevées. Le problème se pose de façon d'autant plus aigüe que les concentrations importées représentent un fort pourcentage des concentrations mesurées ou simulées en aval des zones d'émission du domaine. Or, en Île-de-France, les émissions locales de monoxyde d'azote dépassent de beaucoup les concentrations advectées (ESQUIF 2001) : les concentrations de NO importées dans le domaine atteignent au maximum 5 ppb, que la masse d'air soit d'origine océanique (7 août 1998, POI 5 et 6) ou d'origine continentale (POI 1,2,7 et 8), alors que les concentrations simulées en ville aux mêmes heures dépassent 15 ppb. L'inversion utilisant les concentrations analysées de monoxyde d'azote comme contraintes, les conditions aux limites «propres» ou «polluées» n'ont pas une influence suffisante pour limiter la validité des corrections des émissions.

**La hauteur de couche-limite** Les émissions anthropiques en été en Île-de-France ne dépendent pas de la météorologie, notamment parce qu'il n'y a pas de chauffage en ville. Les émissions et la hauteur de couche-limite réelles peuvent donc être considérées comme des phénomènes indépendants. Cependant, lors de l'optimisation, une mauvaise estima-

Émissions <i>a priori</i>	HCL	
	surestimée	sous-estimée
sous-estimées	$E_{opt}$ surestimées	$E_{opt} \approx E_{ap} \rightarrow$ sous-estimées
correctes	$E_{opt}$ surestimées	$E_{opt}$ sous-estimées
surestimées	$E_{opt} \approx E_{ap} \rightarrow$ surestimées	$E_{opt}$ sous-estimées

TAB. 6.6 – Conséquence sur les émissions optimisées des différentes combinaisons d'erreurs sur le cadastre *a priori* et la hauteur de couche-limite.



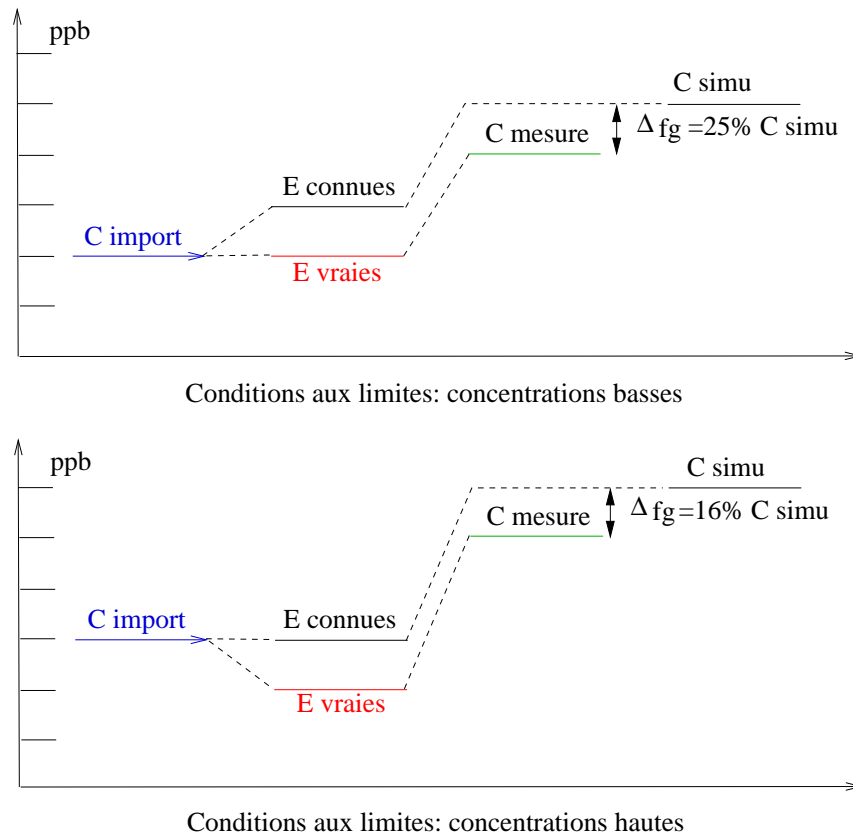


FIG. 6.29 – Illustrations de l'impact des conditions aux limites sur l'information concernant les émissions locales contenue dans l'écart ( $\Delta_{fg}$ ) entre concentrations mesurées et simulées.

tion de la hauteur de couche-limite dans le modèle peut amener une sur-correction des émissions. Les différentes combinaisons d'erreurs sur les émissions *a priori* et la HCL sont résumées dans le Tableau 6.6.

On remarquera que même lorsque le cadastre *a priori* est correct, une HCL trop faible conduit à simuler des concentrations trop élevées et donc à corriger indûment les émissions à la baisse ; inversement, une HCL trop grande amène à des concentrations simulées trop faibles donc à des émissions corrigées à la hausse à tort (Tableau 6.6, ligne du milieu). Enfin, lorsque les erreurs sur les émissions et la HCL sont du même ordre de grandeur, elles peuvent se compenser si la HCL et les émissions sont surestimées ou sous-estimées ensemble. Finalement, quelle que soit l'erreur du cadastre *a priori*, une surestimation (respectivement sous-estimation) de la hauteur de couche-limite amène à des émissions optimisées surestimées (respectivement sous-estimées).

Pour tenter de quantifier l'impact de l'erreur sur la hauteur de couche-limite sur les flux d'émission optimisés, on compare les résultats obtenus pour les deux vendredis étudiés précédemment, le 7 août 1998 et le 16 juillet 1999.

La Figure 6.30 représente les écarts (en %) entre les émissions optimisées obtenues avec la HCL corrigée grâce aux mesures et avec la HCL fournie par l'ECMWF pour ces deux journées. Le 7 août, la HCL fournie par l'ECMWF est augmentée de 50% pendant toute la journée pour approcher les mesures disponibles (représentée en bleu sur la Figure 6.30,

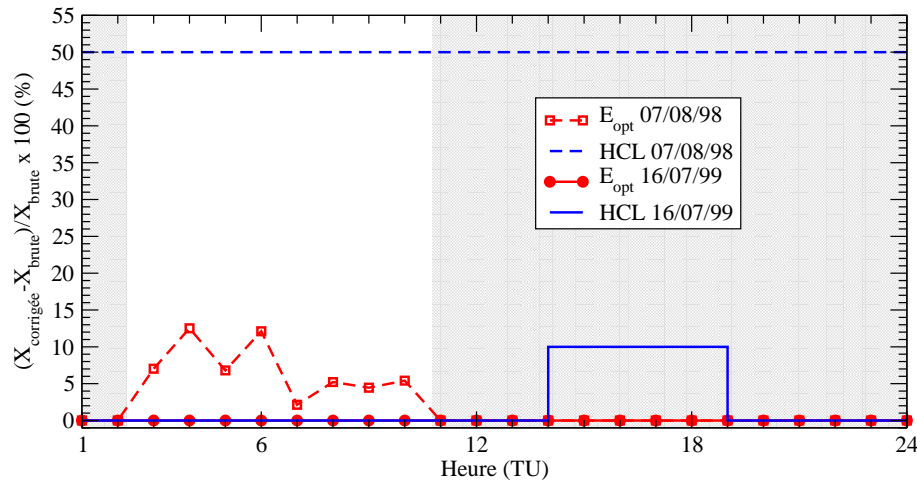


FIG. 6.30 – Différence (en %) entre les profils temporels moyens sur tout le domaine des flux d'émission de NO optimisés avec la hauteur de couche-limite fournie par ECMWF et avec la hauteur de couche-limite corrigée grâce aux mesures. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.

voir page 104). Les flux optimisés obtenus avec la HCL corrigée sont supérieurs de 2% à 7 heures à 12,5% à 4 heures aux flux optimisés avec la HCL brute. Les corrections de HCL effectuées le 16 juillet (augmentation de 10% nécessaire pendant quelques heures seulement, représentée en bleu sur la Figure 6.30, voir page 117) n'ont quasiment pas d'impact sur les flux optimisés : la différence est inférieure à 0,5%.

Le 7 août 1998 est la seule journée, sur l'ensemble des quatorze jours de POI ESQUIF en juillet et août 1998 et 1999, qui présente à la fois une HCL faible (inférieure à 1000 mètres) et une erreur très importante par rapport aux mesures ( $\approx -35\%$ ). L'écart maximal sur les flux optimisés avec et sans correction de la HCL reste 3,5 fois moins grand que la correction apportée à la HCL. Dans le cas d'une HCL plus grande et d'une erreur relative plus faible, comme le 16 juillet 1999, l'influence de la correction de la HCL sur les flux optimisés moyennés sur tout le domaine est négligeable. Finalement, bien que deux cas ne puissent suffire à établir une quantification précise, on peut considérer que la correction de la hauteur de couche-limite est importante pour l'étude de journées spécifiques mais ne représente pas une erreur de premier ordre pour l'inversion d'une longue période comportant peu de situations exceptionnelles en ce qui concerne la HCL.

### Impact global sur les flux d'émission

Dans ce qui suit, pour représenter les résultats de l'optimisation sur la période complète de quatre mois, on utilise la médiane des flux optimisés. En effet, la valeur de cet indicateur n'est pas influencée, comme la moyenne, par les valeurs extrêmes, qui peuvent correspondre à des intensités aberrantes dues à une sur-correction des émissions.

**Répartition temporelle** On examine dans un premier temps l'ensemble des jours optimisés afin de quantifier la variabilité des émissions optimisées. Le cadastre *a priori* présente une variabilité due au fait qu'il existe six types de profils différents, représentés

sur la Figure 6.31 : un pour chacun des trois types de jour (semaine, samedi, «dimanche ou jour férié»), avec des intensités différentes en juillet et en août (20% de moins, voir page 109).

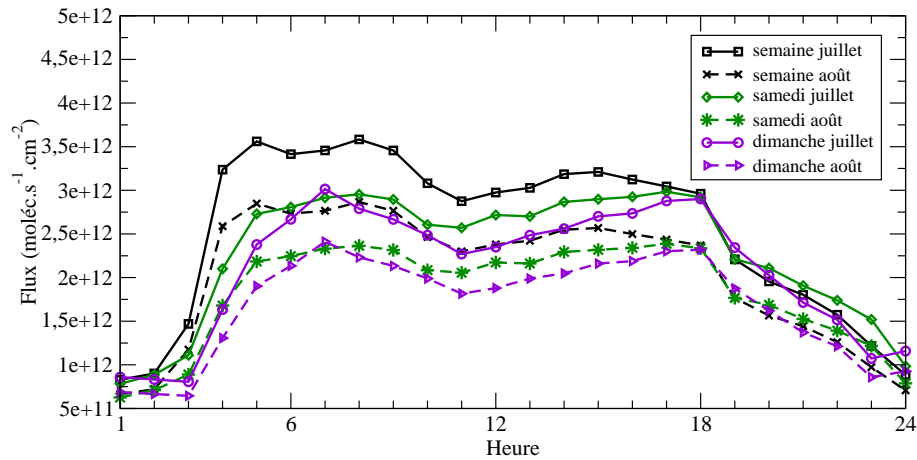


FIG. 6.31 – *Profils temporels a priori des flux d'émission de NO moyennés sur tout le domaine, selon les différents types de jour pour les mois de juillet et août 1998 et 1999. En été, TU = heure légale - 2 heures.*

La Figure 6.32 représente donc la médiane des profils temporels des 124 jours optimisés avec les percentiles 50 et 80 et les deux profils temporels correspondant aux variations extrêmes du cadastre *a priori*.

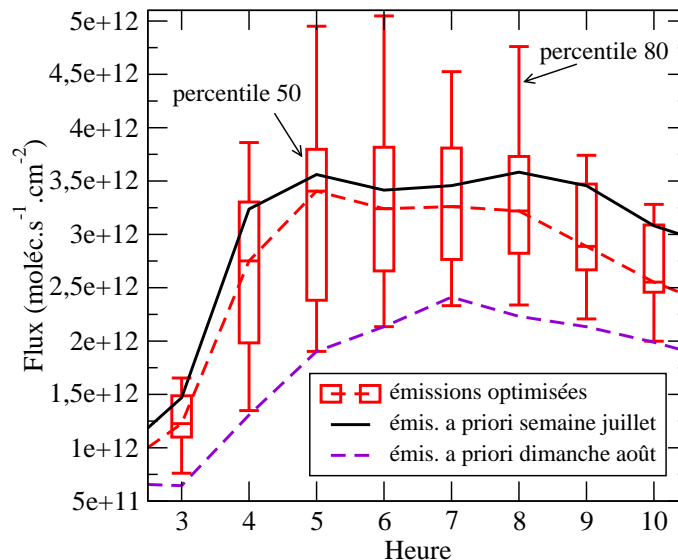


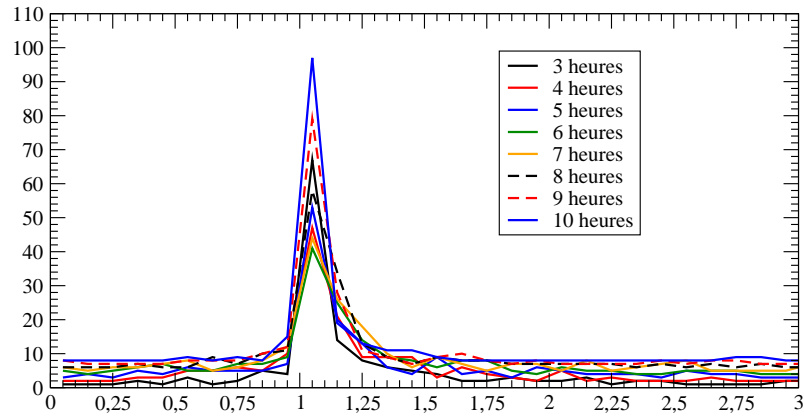
FIG. 6.32 – *Profil temporel médian des flux d'émission de NO moyennés sur tout le domaine, obtenu sur les périodes du 1er juillet au 31 août 1998 et 1999. En été, TU = heure légale - 2 heures.*

La moitié supérieure du percentile 80 indique une grande dispersion des flux moyens optimisés, notamment de 4 à 8 heures. Comme indiqué plus haut, parmi les flux optimisés très éloignés de la médiane se trouvent probablement des résultats dus à une sur-correction des émissions effectuée à cause d'une mauvaise estimation de la HCL. Cependant, cela n'explique pas la dissymétrie, qui apparaît à 5 et 6 heures dans les intensités : la limite supérieure du percentile 80 correspond à des flux atteignant +59% de la médiane. Un examen des profils temporels journaliers montre que, lorsque la HCL est faible le matin avant 6 heures (inférieure à 300 mètres) puis double à 7 heures, le nombre de flux supérieurs au percentile 80 est deux fois plus grand que le nombre de flux inférieurs. La variation rapide de la HCL semble donc être mal prise en compte par l'inverse qui a tendance à augmenter de façon disproportionnée les émissions pour compenser la dilution brutale des concentrations.

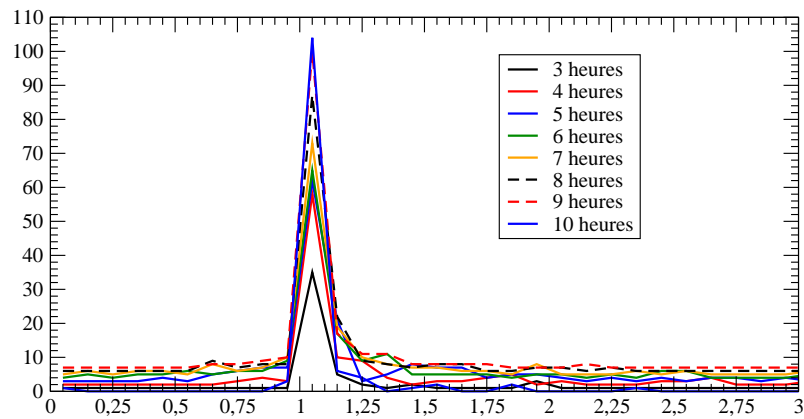
Entre 3 et 10 heures, le percentile 50 du profil médian optimisé correspond à des flux d'émission variant de -51 à +56% autour de la médiane. Cette variabilité est donc du même ordre de grandeur que l'écart d'intensité *a priori* de 20% entre le mois d'août et le mois de juillet d'une part et que la différence de 35% dans la masse émise de 3 à 10 heures entre les jours de la semaine et le type «dimanche ou jour férié» d'autre part (Figure 6.31). La variabilité de l'ensemble des flux optimisés paraît donc réaliste. On peut comparer dans ce qui suit les résultats obtenus pour le cadastre optimisé au cadastre *a priori* pour les différents types de jour utilisés.

**Répartition spatiale** L'étude du 7 août 1998 et du week-end du 16 au 18 juillet 1999 a montré que la répartition spatiale des corrections n'était pas homogène mais dépendait du lieu et de l'heure (voir pages 107 et 118). Parmi l'ensemble des coefficients de correction appliqués à une même maille au cours de la période optimisée, certaines valeurs sont obtenues avec une fréquence très supérieure aux autres. La Figure 6.33 représente les distributions en fréquence des coefficients de correction obtenus pour la maille rurale de Paris Sud-Ouest et pour la maille périphérique de Tremblay, avec des classes de 0,05. Dans la maille urbaine (Figure 6.33 (a)), à chaque heure optimisée, une seule classe dépasse 50 occurrences sur un total de 124. Dans la maille périphérique, la distribution présente également un pic de fréquence pour une seule classe, qui dépasse 35 occurrences (Figure 6.33 (b)). On choisit donc d'utiliser le coefficient le plus fréquemment obtenu dans chaque maille pour représenter la répartition spatiale des corrections.

La Figure 6.34 indique, pour chaque heure inversée dans chaque maille optimisée, la correction la plus fréquente sur l'ensemble des jours optimisés ; pour des raisons de lisibilité, toutes les classes ne sont pas représentées individuellement. On remarquera tout d'abord que, quelle que soit l'heure considérée, les corrections le plus souvent effectuées consistent en l'augmentation des émissions de  $\text{NO}_x$ , ce qui est cohérent avec la sous-estimation des concentrations matinales de NO, au moins en ville indiquée par les comparaisons statistiques (voir page 93). Entre 6 et 8 heures ainsi qu'à 10 heures, ces corrections sont réparties différemment selon l'intensité des flux. En effet, les flux les plus intenses du centre du domaine sont le plus fréquemment diminués par l'optimisation (jusqu'à -5% par rapport aux valeurs *a priori*, coefficients compris entre 0,95 et 1) alors que la majorité des flux périphériques moins intenses sont augmentés (jusqu'à +5% par rapport aux valeurs *a priori*, coefficients compris entre 1 et 1,05). De 3 à 5 heures et à 9 heures, la répartition spatiale est plus homogène, la plupart des flux étant fréquemment multipliée par des coefficients compris entre 1 et 1,05.



(a) Paris Sud-Ouest



(b) Tremblay

FIG. 6.33 – Distribution en fréquence (pour des classes de 0,05) des coefficients de correction obtenus à chaque heure inversée sur la période du 1er juillet au 31 août 1998 et 1999. Pour plus de lisibilité, la courbe indiquant la distribution de chaque heure est décalée vers le haut d'une unité de plus que la précédente.

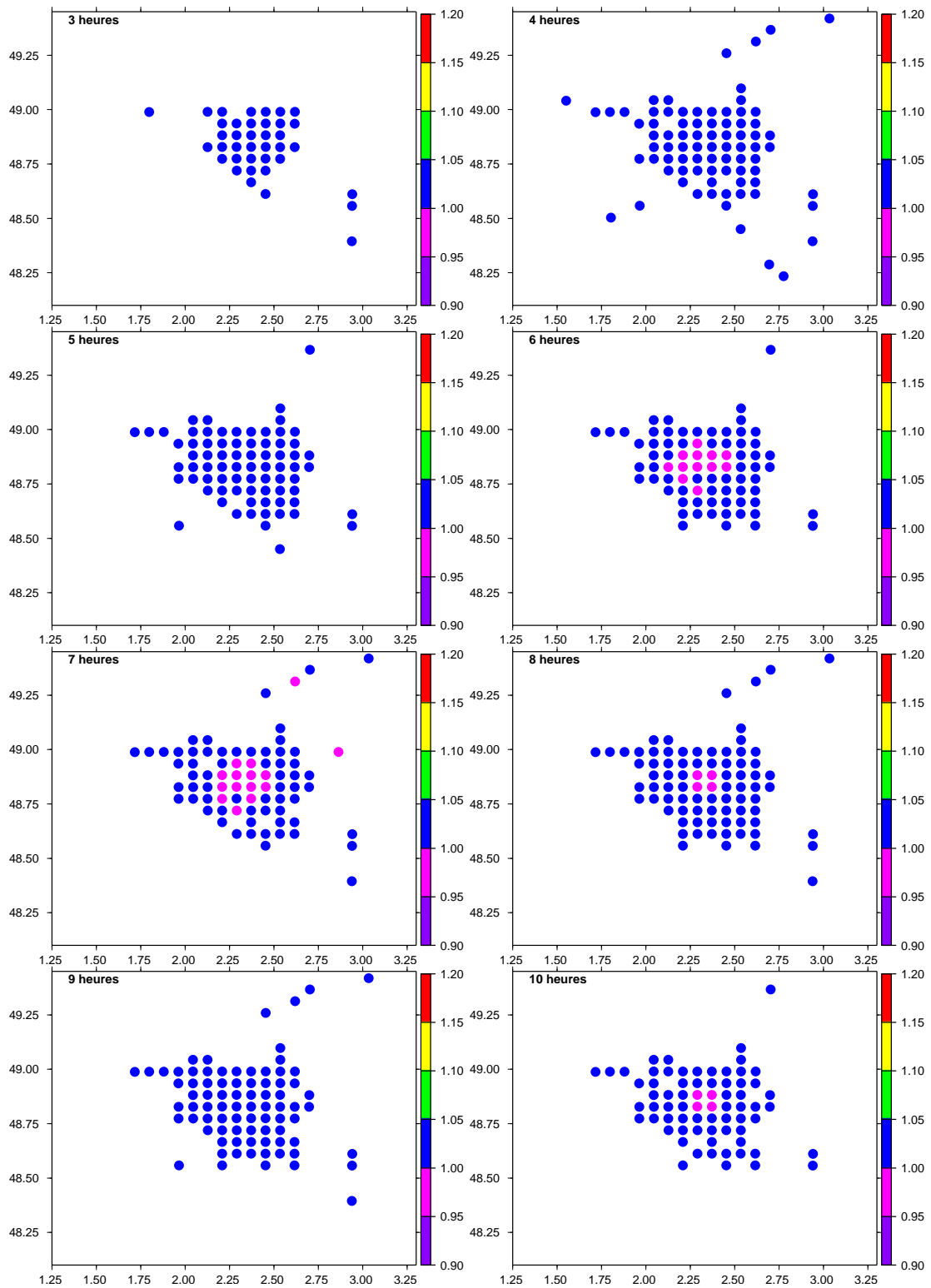


FIG. 6.34 – Représentation spatiale des corrections les plus fréquemment appliquées aux flux d'émission de  $\text{NO}_x$  sur les périodes du 1er juillet au 31 août 1998 et 1999. Classes de 0,05 non représentées individuellement. En été, TU = heure légale - 2 heures.

### Impact sur les différents profils temporels

**Mois de juillet et août** Comme indiqué plus haut, les émissions *a priori* du mois d'août sont, à toute heure, inférieures de 20% aux émissions du mois de juillet. Les profils temporels médians des flux optimisés aux mois de juillet et août, représentés sur la Figure 6.35, montrent que cette différence est presque conservée après optimisation : elle varie entre -19,2% à 8 heures et -22,5% à 6 heures.

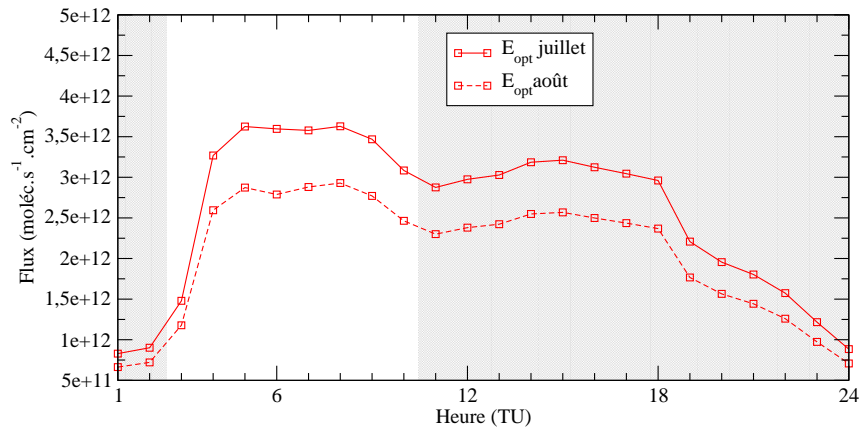


FIG. 6.35 – Profils temporels médians des flux d'émission de NO optimisés moyennés sur tout le domaine, obtenu sur les mois de juillet et août 1998 et 1999. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.

**Semaine** Par rapport au profil temporel *a priori*, l'intensité des émissions optimisées est augmentée de 9,4 à 17,7% entre 4 et 10 heures (Figure 6.36). Cependant, l'allure du profil est conservée, avec deux pics, à 5 et 8 heures. Au total, la masse émise est augmentée de 13,1%. L'intensité plus élevée des flux d'émission optimisés est un résultat

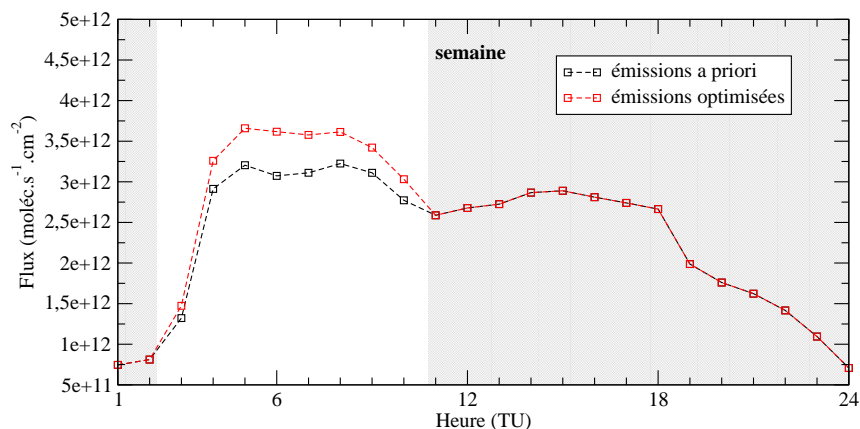


FIG. 6.36 – Profil temporel médian sur tout le domaine des flux d'émission de NO pour le type «semaine» obtenu sur les périodes du 1er juillet au 31 août 1998 et 1999. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.

attendu. En effet, la comparaison des concentrations de monoxyde d'azote simulées *a priori* et mesurées avait mis en évidence la sous-estimation des concentrations matinales de NO, au moins en ville (voir page 93). La conservation de l'allure du profil temporel est cohérente avec les résultats de l'étude des corrélations qui avait indiqué que la variabilité et le phasage temporel des concentrations de monoxyde d'azote étaient corrects, au moins en milieu urbain (voir page 95).

**Samedi et dimanche/jour férié** Le samedi (Figure 6.37 en haut), l'intensité des émissions optimisées est réduite, entre 4 et 10 heures, de 1,9 à 16,3% par rapport au cadastre *a priori*. Les réductions les plus importantes sont obtenues à 5 et 6 heures (13,2 et 16,3%). Au total, la masse émise est inférieure de 7,7% pour le profil optimisé par rapport au profil *a priori*. Le profil optimisé comporte un pic d'émissions entre 7 et 9 heures (flux supérieurs à  $2,5 \cdot 10^{12}$  moléc.cm<sup>-2</sup>.s<sup>-1</sup>) alors que les flux *a priori* sont quasiment constants entre 5 et 9 heures (flux compris entre 2,7 et  $2,9 \cdot 10^{12}$  moléc.cm<sup>-2</sup>.s<sup>-1</sup>). On peut évoquer la possibilité que ce pic corresponde au trafic de «loisir» du samedi matin, plus tardif et moins intense que le trafic dû aux déplacements de travail en semaine.

Le dimanche (Figure 6.37 en bas), l'impact de l'optimisation est plus faible que le samedi : les flux optimisés sont augmentés de 15,6 et 10,5% à 8 et 9 heures par rapport aux flux *a priori* et différent de ceux-ci de moins de 5% de 4 à 7 heures. La masse émise entre 3 et 10 heures est augmentée de moins de 7% après optimisation. Le profil médian du type «dimanche ou jour férié» a été élaboré à partir de 18 dimanches plus les 14 juillet 1999 et 1998 et le samedi 15 août 1998. Un sixième des flux optimisés concernent donc des jours très particuliers pour le trafic. Par exemple, le mardi 14 juillet 1998 a pu donner lieu à des retours de «pont» (depuis le samedi 11), bien que ceux-ci influencent en principe les émissions de la soirée. Finalement, l'allure des profils optimisés diffère peu de celle des profils *a priori* qui étaient cependant supposés comporter une grande incertitude (voir le 17 et 18 juillet 1999 page 113).

**Les différents jours de la semaine** Un unique profil du type «jour de la semaine» a été utilisé lors de l'élaboration du cadastre *a priori*. Or, le trafic comptant pour la majorité des émissions de NO<sub>x</sub> (53% des émissions totales dans le cadastre *a priori*), on peut supposer que les profils temporels des différents jours de la semaine présentent des caractéristiques particulières.

Pour différencier chacun des cinq jours de la semaine dans le cadastre optimisé, il faut comparer le profil médian de chaque jour au profil médian de toute la semaine. Afin de séparer la variabilité due aux différences d'intensité entre les mois de la variabilité éventuelle due aux différences entre les jours, les médianes ont été calculées séparément pour juillet et août. Au total, on compte dix-huit lundis (8 en juillet, 10 en août), jeudis (10 en juillet, 8 en août) et vendredis (10 vendredis en juillet, 8 en août), seize mardis (7 en juillet, le 14 juillet 1998 étant un mardi, 9 en août) et seize mercredis (8 en juillet, le 14 juillet 1999 étant un mercredi, 8 en août).

Pour le mois de juillet, les profils temporels médians des lundi, mardi, jeudi et vendredi sont très semblables au profil temporel médian de la semaine : les écarts sont inférieurs à 3,5% à toute heure, l'écart maximal étant de 3,4% le jeudi à 4 heures (Figure 6.38). Le profil temporel du mercredi est également très proche du profil typique de la semaine (différences inférieures à 2,5%) sauf à 7 heures où l'écart atteint +5,6% (Figure 6.38).



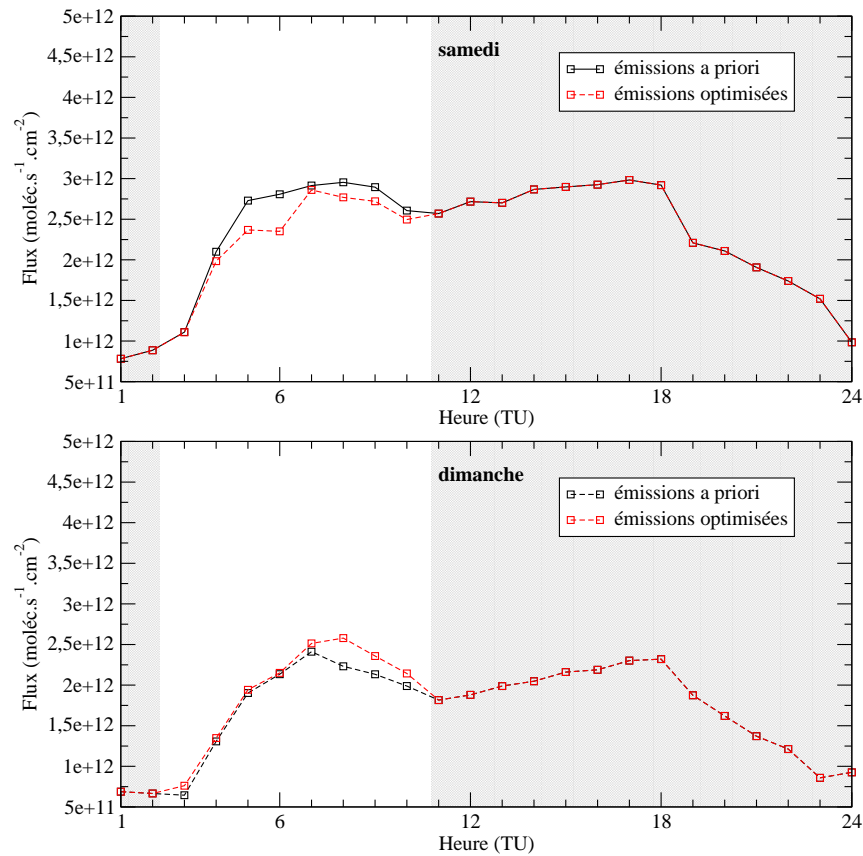


FIG. 6.37 – Profils temporels médians sur tout le domaine des flux d'émission de NO pour les types «samedi» et «dimanche ou jour férié» obtenus sur les périodes du 1er juillet au 31 août 1998 et 1999. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.

Les flux d'émission du mercredi ne sont donc pas quasiment constants de 5 à 8 heures (à environ  $3,7 \text{ moléc.cm}^{-2}.\text{s}^{-1}$ ), comme avec le profil hebdomadaire, mais forment un pic à plus de  $3,9 \text{ moléc.cm}^{-2}.\text{s}^{-1}$  à 6 heures.

Au mois d'août, les profils temporels médians des mardi, jeudi et vendredi sont très semblables au profil temporel médian de la semaine : les écarts sont inférieurs à 7% à toute heure, l'écart maximal étant de -6,9% le mardi à 6 heures (Figure 6.38). Le profil temporel du mercredi conserve l'allure du profil hebdomadaire mais avec une intensité, entre 4 et 8 heures, supérieure de 0,8 à 8,3%. Le profil temporel du lundi est également très proche du profil typique de la semaine (différences inférieures à 3,2%) sauf entre 5 et 6 heures où l'écart atteint +11,2 et +18,4% (Figure 6.38). Les flux d'émission du lundi présentent donc, à 5 heures, un pic à près de  $3,7 \text{ moléc.cm}^{-2}.\text{s}^{-1}$ . On peut en proposer une explication en terme de trafic de retour de longs week-end. En effet, dans l'ensemble des 10 lundis du mois d'août se trouvent quatre lundis susceptibles d'être des jours de trafic dû aux vacanciers (les lundis 2 et 30 août 1999 et les lundis 3 et 31 août 1999) ainsi que les deux lundis 16 août 1999 et 17 août 1998 clôturant les week-ends du 15 août.

Finalement, les cinq jours de la semaine ne présentent pas de caractéristiques moyennes très marquées par rapport aux profils hebdomadaires. Cependant, depuis 1999, on peut

s'attendre à ce que les habitudes de déplacement de la population aient été modifiées par la loi dite «des 35 heures». Cela pourrait faire apparaître des différences systématiques plus importantes entre le profil typique de la semaine et les lundis et vendredis par exemple.

### 6.6.3 Synthèse des résultats

L'optimisation des émissions de  $\text{NO}_x$  matinales sur les mois de juillet et août 1998 et 1999 a permis d'obtenir des concentrations d'ozone en meilleur accord avec les mesures. Le biais et l'erreur moyenne ont été diminués de 1,6 et 2,3% et la corrélation a été améliorée de 1,5%. Après avoir comparé les 7 août 1998 et 16 juillet 1999, on considère que la correction de la hauteur de couche-limite, importante pour l'étude de journées spécifiques (c'est-à-dire comportant des situations météorologiques extrêmes), n'est pas essentielle pour l'inversion d'une longue période comportant peu de situations exceptionnelles. L'examen de la médiane de l'ensemble des flux optimisés montre que la variabilité du percentile 50, regroupant donc la moitié des jours inversés, est comparable à celle qui a été introduite dans le cadastre *a priori* (différence entre les mois et les types de jour). Les corrections les plus fréquemment effectuées sont réparties spatialement selon l'intensité des flux d'émission. Après optimisation, les flux *a priori* les plus intenses du centre du domaine sont le plus souvent diminués de 65% ou plus alors que l'intensité de la majorité des flux moins intenses en périphérie est modifiés de moins de 55%. Pour tous les types de jour (semaine, samedi et «dimanche ou jour férié»), les écarts entre profils optimisés et *a priori* sont inférieurs à 15%. Cette valeur rejoint les résultats obtenus lors de l'évaluation du cadastre d'émission effectuée grâce aux mesures aéroportées au cours de la campagne ESQUIF (ESQUIF 2001) : l'erreur globale sur les émissions de  $\text{NO}_x$  avait alors été estimée à 15% au maximum. Enfin, la comparaison entre le profil obtenu pour toute la semaine et celui de chacun des cinq jours ne fait apparaître de caractéristique particulière que pour les lundis et les mercredis du mois d'août. Cependant, on peut supposer que les caractéristiques du trafic des différents jours de la semaine ont été modifiées depuis 1999.

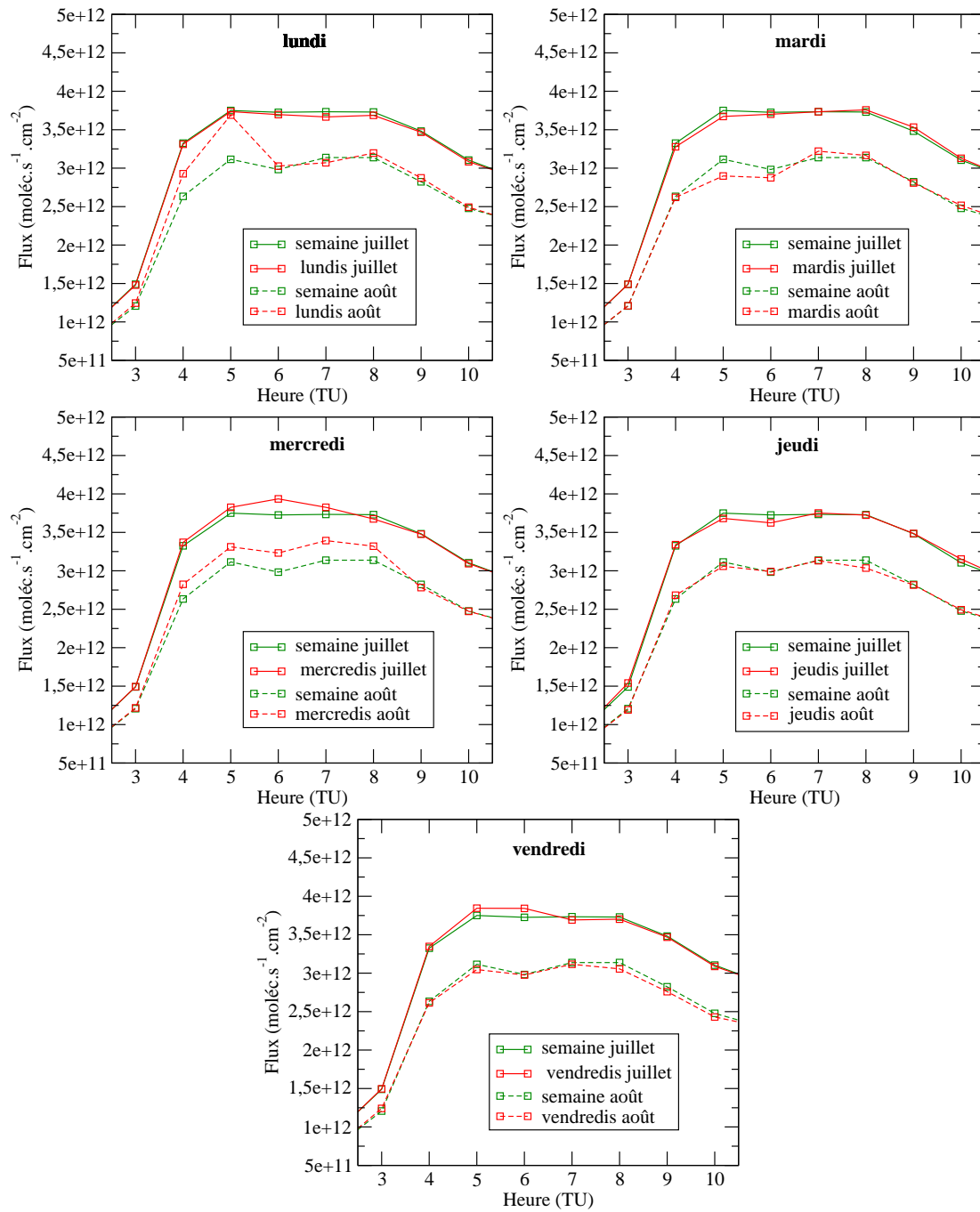


FIG. 6.38 – Profils temporels médian sur tout le domaine des flux d'émission de NO optimisés pour les jours de type «semaine» obtenus sur les mois de juillet et août 1998 et 1999. En été, TU = heure légale - 2 heures.

## 6.7 Conclusion

L'Île-de-France constitue un bon cadre aux études d'inversion des émissions. En effet, dans cette région, les champs dynamiques sont bien reproduits par le modèle. De plus, le réseau de mesures AIRPARIF y est bien développé.

Après que l'influence des émissions tri-dimensionnelles dues aux avions de ligne a été quantifiée et trouvée faible, il a été décidé de n'inverser que les émissions anthropiques surfaciques.

Une étude climatologique de comparaison entre concentrations d'ozone et de monoxyde d'azote simulées et mesurées au cours des étés 1998 et 1999 a permis de déterminer les points faibles des simulations de CHIMERE. Les indicateurs statistiques (biais, moyenne quadratique et corrélation) ont indiqué que les concentrations d'ozone lors des pics étaient bien reproduites par le modèle (biais inférieur à 10 ppb pour des concentrations supérieures à 100 ppb) malgré la sous-estimation des concentrations matinales de monoxyde d'azote en ville (biais supérieur à -10 ppb). Ce défaut peut être relié soit à une mauvaise estimation de l'intensité du mélange vertical soit à une sous-estimation des flux d'émission de  $\text{NO}_x$ .

Une étude simplifiée de l'impact du coefficient de mélange turbulent et de la hauteur de couche-limite sur les concentrations lors de la POI 2 d'ESQUIF a indiqué que la paramétrisation du mélange vertical pouvait être considérée comme correcte. La comparaison des flux d'émissions optimisés obtenus avec et sans correction de la hauteur de couche-limite lors de deux journées des POI 2 et 6 (les vendredis 7 août 1998 et 16 juillet 1999) ont indiqué qu'il était nécessaire de débiaiser la hauteur de couche-limite pour l'étude de journées très particulières, comportant des situations météorologiques extrêmes mais pas pour l'inversion d'une longue période comportant peu de situations exceptionnelles.

L'étude de cas réels effectuée ici comporte deux volets : d'une part, l'inversion de cas particuliers de pollution, d'autre part, l'inversion d'une climatologie composée des deux étés 1998 et 1999.

Dans tous les cas, on inverse les émissions matinales (de 3 à 10 heures) de  $\text{NO}_x$ , ce qui modifie donc le rapport COV sur  $\text{NO}_x$  à l'émission. Parmi les hypothèses de travail retenues, trois sont particulièrement importantes : (i) les paramétrisations du modèle sont assez réalistes, (ii) sauf pour la HCL de certaines journées, la météorologie est considérée comme assez réaliste pour n'avoir pas besoin d'être inversée ou débiaisée et (iii) les sources d'émission sont bien placées dans le cadastre *a priori*, l'optimisation porte sur leur intensité.

**Deux cas de pollution dite locale**, dans lesquels les émissions du domaine jouent donc un rôle prépondérant, constituent les POI 2 et 6 de la campagne ESQUIF. Les émissions matinales de  $\text{NO}_x$  du vendredi 7 août 1998 et des vendredis 16, samedi 17 et dimanche 18 juillet 1999 ont été inversées après que la hauteur de couche-limite a été corrigée en fonction des mesures obtenues par les instruments supplémentaires.

Les résultats obtenus pour les deux vendredis sont très différents. En effet, le profil temporel des émissions du 7 août 1998 est modifié de telle sorte que la durée du pic matinal en ville est réduite de cinq heures *a priori* à deux heures après optimisation alors que le profil temporel du 16 juillet 1999 est conservé. Ces modifications peuvent être reliées au fait que le cadastre *a priori* a été élaboré sur la base du mois de juillet, les émissions du mois d'août étant obtenues en diminuant tous les flux de 20% mais en suivant les mêmes

variations diurnes. Les modifications apportées par l'optimisation le 7 août peuvent donc correspondre à une différence réelle dans le trafic entre les mois de juillet et août.

Les profils temporels de type «samedi» et «dimanche et jour férié» utilisés pour l'élaboration du cadastre *a priori* comportent une grande incertitude. Le samedi 17 juillet 1999 présente finalement des résultats très contrastés selon les lieux. En ville, le pic matinal d'émission de  $\text{NO}_x$  est réduit en durée (de cinq heures *a priori* à deux heures après optimisation) mais augmenté de 25% en intensité. À l'extérieur de la ville, le pic matinal d'émission disparaît du cadastre optimisé. Le dimanche 18 juillet, le cadastre optimisé permet de rectifier une erreur importante en éliminant un pic d'émission simulé en ville mais non observé. Il comporte néanmoins, pour les flux moins intenses, des modifications peu réalistes

La répartition spatiale des corrections a montré que les résultats obtenus pour les flux de la ville, c'est-à-dire là où sont concentrées de nombreuses mesures, étaient plus réalistes que ceux obtenus pour les flux périphériques, situés dans des zones où le réseau de mesure est moins dense. L'utilisation du krigeage pour répartir au mieux l'information disponible dans les concentrations mesurées ne permet donc pas de surmonter toutes les limitations dues à la quantité limitée d'information.

Finalement, l'étude de deux cas particulièrement bien documentés a montré que la méthodologie d'inversion était applicable à des cas réels. Une étude d'inversion à plus long terme a donc été menée pour obtenir des résultats généralisables à d'autres périodes.

**L'inversion à long terme** a porté sur **les mois de juillet et août 1998 et 1999**. Sur l'ensemble des profils d'émission optimisés, la moitié présente une variabilité inter-journalière du même ordre de grandeur que les écarts utilisés lors de l'élaboration du cadastre *a priori* pour reproduire les différences entre les mois (juillet et août) ou les jours (semaine par rapport au dimanche ou jour férié). La répartition spatiale des corrections les plus fréquentes sur toute la période inversée confirme les différences entre les flux intenses du centre du domaine, donc l'intensité est souvent corrigée de plus de 65%, et les flux périphériques moins intenses et modifiés dans des proportions moindres.

La comparaison des profils optimisés et *a priori* pour les trois types de jour (semaine, samedi et «dimanche ou jour férié») montre des écarts inférieurs à 15%, ce qui correspond à l'estimation de l'erreur globale sur les émissions de  $\text{NO}_x$  effectuée lors de la campagne ESQUIF.

Enfin, seuls les profils temporels des lundis et mercredis du mois d'août présentent des caractéristique particulières différentes de celles du profil hebdomadaire typique. Cependant, ces résultats ne sont probablement valables que pour les années antérieures à 1999.

# Chapitre 7

## Conclusions et perspectives

### 7.1 Conclusions

#### Rappel du cadre du travail

Ce travail avait pour objectif d'élaborer une méthode d'inversion des émissions anthropiques de  $\text{NO}_x$  et de COV à l'échelle régionale, utilisant les mesures de routine des réseaux ( $\text{O}_3$  et  $\text{NO}_x$ ). Ces dernières étant disponibles dans de nombreuses régions et pour de longues périodes, une telle méthode permettrait (i) d'estimer a posteriori les émissions nécessaires à une meilleure simulation des champs de concentrations lors d'épisodes passés variés, (ii) d'envisager une utilisation de l'inversion pour l'amélioration des prévisions. De plus, une meilleure connaissance des émissions passées et actuelles peut offrir une base plus sûre à l'élaboration de scénarios prospectifs.

Bien que les méthodes de la famille de l'assimilation soient déjà utilisées dans de nombreux domaines et notamment en météorologie et en chimie de l'atmosphère à l'échelle globale, les études d'inversion des émissions à l'échelle régionale sont actuellement peu nombreuses. Il était donc nécessaire d'évaluer ce que les méthodes inverses peuvent apporter à la simulation de la pollution photo-oxydante à l'échelle régionale.

#### Méthodologie élaborée

La principale limitation des méthodes inverses réside dans la quantité d'information nécessaire. C'est pourquoi le nombre limité de données d'observation disponibles a constitué le principal problème à résoudre. Deux stratégies ont donc été mises en œuvre simultanément : d'une part, augmenter le nombre de contraintes et leur «adéquation» par rapport au problème à inverser, en utilisant les informations contenues dans l'espace des concentrations, et d'autre part, diminuer la dimension du problème, en utilisant les informations contenues dans l'espace des émissions.

La structure du champ de concentrations simulé *a priori* par CHIMERE ayant été validée, une technique de krigeage a été mise en place pour utiliser les informations contenues dans l'espace des concentrations. Les contraintes constituées par le champ de concentrations analysé prennent en compte à la fois les informations fournies par les mesures et par le champ simulé. Elles représentent la meilleure information possible dans l'espace des concentrations (notamment par rapport à la représentativité du maillage).

Dans l'espace des paramètres, une technique d'agrégation spatiale dynamique a été élaborée : des zones regroupent les paramètres qui sont à corriger dans la même proportion.

La dimension du problème est ainsi réduite du nombre total de paramètres au nombre de zones défini. Dans le cas des flux d'émission en Île-de-France, on regroupe des flux d'émissions d'intensités proches. De plus, la dynamique de la situation est prise en compte afin de définir des zones dans lesquelles le signe de la sensibilité du coût à ces émissions est le même.

La méthode de l'adjoint a été choisie pour notre travail parce qu'elle permet de prendre en compte facilement l'évolution temporelle des concentrations et des émissions. L'inverse du modèle CHIMERE a donc été développé à partir du modèle direct version V200211K, de son adjoint (qui a été revu et modifié) et de l'optimiseur N1QN3.

### **Validation et limites de l'outil**

La méthodologie élaborée ne comporte aucune hypothèse limitant sa validité a priori. elle peut donc en principe être utilisée pour toutes les périodes de l'année, pour toutes les espèces pour lesquelles on dispose d'une technique de krigeage adaptée. Cependant, la réalisation de l'agrégation spatiale dépend des caractéristiques dynamiques du domaine considéré.

Pour évaluer les capacités de l'inversion, une procédure permettant de générer et d'étudier des cas académiques a été mise en place.

Si toute l'information nécessaire est disponible, on vérifie que la précision de l'inverse est égale à la précision de la machine.

La méthode avec krigeage a été comparée à une méthode utilisant des mesures seules et à une méthode utilisant une matrice «de background». Les contraintes fournies par le krigeage permettent d'obtenir des corrections dont la répartition spatiale et temporelle est plus homogène et cohérente pour l'ensemble du cadastre.

Cependant, le krigeage ne permet pas de créer de l'information. Il est donc nécessaire de choisir comme contrainte une espèce pour laquelle sont disponibles suffisamment de données d'observation contenant de l'information sur les émissions à inverser. En Île-de-France par exemple, on ne peut utiliser les concentrations d'ozone pour inverser les émissions de  $\text{NO}_x$  matinales car on ne dispose en général que d'une seule observation dans le panache de l'après-midi.

### **Étude en Île-de-France**

Les premières études d'inversion ont porté sur le cadastre élaboré par AIRPARIF pour l'Île-de-France. En effet, cette région présente une topographie simple (relief faible, éloignement des côtes) et le réseau de mesure AIRPARIF est suffisamment dense et complet pour le développement et la validation de notre méthode. Enfin, plusieurs cas de pollution ont été bien documentés et étudiés en détail lors de la campagne de mesure ESQUIF des étés 1998 et 1999.

Dans un premier temps, les écarts entre concentrations mesurées et simulées ont été examinés afin de déterminer les éventuels problèmes rencontrés par le modèle.

Une étude climatologique de comparaison entre concentrations d'ozone et de monoxyde d'azote simulées et mesurées au cours des étés 1998 et 1999 a permis de déterminer les points faibles des simulations de CHIMERE. Les indicateurs statistiques calculés montrent que le cycle diurne des concentrations d'ozone et plus particulièrement les valeurs atteintes lors des pics étaient bien reproduits par le modèle (biais inférieur à 10 ppb

pour des concentrations supérieures à 100 ppb). Cependant, les concentrations matinales de monoxyde d'azote en ville sont sous-estimées.

De façon plus ponctuelle, deux cas de pollution dite locale, dans lesquels les émissions du domaine jouent donc un rôle prépondérant, ont été étudiés lors de la campagne ESQUIF. La journée du 7 août 1998 (POI 2 d'ESQUIF) présente un épisode local de pollution photo-oxydante au cours duquel les concentrations d'ozone simulées dans le panache l'après-midi sont sous-estimées alors que les concentrations de  $\text{NO}_x$  matinales en ville sont surestimées.

Les journées du vendredi 16, samedi 17 et dimanche 18 juillet 1999 (POI 6 d'ESQUIF), ont été choisies parce qu'elles correspondent aux trois types de jours utilisés pour l'élaboration du cadastre (semaine, samedi et dimanche ou jour férié). Pour les deux derniers types, les profils temporels diurnes des émissions comportent une grande incertitude. Pendant ces trois jours, les concentrations de monoxyde d'azote simulées le matin en ville et dans les zones péri-urbaines sont très surestimées par rapport aux mesures, particulièrement le samedi et le dimanche.

Dans un deuxième temps, la détermination des paramètres ayant une influence prépondérante sur les écarts constatés a été effectuée. Les mauvaises estimations des concentrations de monoxyde d'azote simulées le matin en ville peuvent être reliées soit à une mauvaise estimation de l'intensité du mélange vertical (qui dépend de la paramétrisation du coefficient de mélange turbulent et de la hauteur de couche-limite), soit à une sous-estimation des flux d'émission de  $\text{NO}_x$ . Une étude de sensibilité a permis de constater que la paramétrisation du coefficient de mélange turbulent ne pouvait expliquer tous les écarts entre concentrations mesurées et simulées. Les bases physiques de cette paramétrisation ayant été validée par ailleurs, il a été décidé de ne pas la modifier. Par contre, la correction de la hauteur de couche-limite fournie par l'ECMWF grâce aux mesures effectuées lors d'ESQUIF, permet d'expliquer une grande part de l'écart entre concentrations de NO mesurées et simulées le matin en ville (près de 40% dans le cas étudié).

Enfin, les trois cas retenus ont été inversés : le 7 août 1998, les 16, 17 et 18 juillet 1999 et les deux étés (mois de juillet et août) 1998 et 1999. Le krigeage des concentrations est effectué grâce aux mesures du réseau AIRPARIF : les contraintes sont les analyses de NO effectuées de 4 à 16 heures. Les flux de  $\text{NO}_x$  les plus intenses (supérieurs à  $5 \times 10^{11} \text{moléc.cm}^{-2}.\text{s}^{-1}$ ) sont optimisés de 3 à 10 heures du matin.

### Synthèse des résultats

L'optimisation permet de corriger certains écarts très importants entre concentrations mesurées et simulées. Par exemple, le dimanche 18 juillet, un pic d'émission de  $\text{NO}_x$  était simulé en ville avec le cadastre *a priori* mais il n'était pas mesuré ; avec le cadastre optimisé, ce pic n'est plus simulé.

Les résultats de l'optimisation d'une journée donnée diffèrent selon les lieux. D'une façon générale, le réalisme des corrections obtenues est plus satisfaisant pour les flux intenses, en ville, c'est-à-dire là où sont concentrées de nombreuses mesures, que pour les flux périphériques, moins intenses et situés dans des zones où le réseau de mesure est moins dense. Le dimanche 18 juillet 1999, par exemple, le cadastre optimisé comporte, pour les flux des mailles rurales, des modifications peu réalistes (un pic d'émission très intense à 7 heures). Malgré l'utilisation du krigeage, qui permet de répartir au mieux dans tout le domaine l'information disponible dans les observations, le niveau de fiabilité des résultats



diminue donc avec la densité du réseau de mesure. Dans le cas des émissions de  $\text{NO}_x$  optimisées grâce aux concentrations de monoxyde d'azote, les résultats présentant le plus haut niveau de confiance sont finalement ceux qui concernent les flux les plus intenses, qui contribuent le plus aux épisodes de pollution en Île-de-France.

L'optimisation des émissions sur la période regroupant les mois de juillet et août 1998 et 1999 a été effectuée avec les mesures du réseau de surveillance de la qualité de l'air. Elle a permis d'obtenir des profils temporels présentant une variabilité inter-journalière du même ordre de grandeur que les écarts utilisés lors de l'élaboration du cadastre *a priori* pour reproduire les différences entre les mois (juillet et août) ou les jours (semaine par rapport au dimanche ou jour férié). Les écarts entre profils optimisés et *a priori* sont inférieurs à 15%, ce qui correspond à l'estimation de l'erreur globale sur les émissions de  $\text{NO}_x$  effectuée lors de la campagne ESQUIF. De plus, les émissions optimisées permettent d'améliorer la simulation des concentrations d'ozone sur le long terme (biais et erreur moyenne diminués par rapport aux valeurs *a priori*, corrélation augmentée). Finalement, les corrections apportées à la masse moyenne émise dans l'ensemble du domaine et aux profils temporels correspondant sont donc réalistes.

## 7.2 Perspectives

### Validation extérieure des cadastres optimisés

Ce travail a principalement porté sur l'élaboration et la validation de la méthodologie d'inversion à l'échelle locale en utilisant les mesures de routine du réseau. Après les premières études réalisées en Île-de-France pendant les étés 1998 et 1999, la question de la validation extérieure des cadastres optimisés reste donc à discuter. En effet, si les émissions optimisées paraissent cohérentes et permettent de réduire l'écart entre concentrations simulées et mesurées, elles peuvent néanmoins ne pas correspondre à une réalité physique et dépendre du modèle utilisé. L'utilisation des cadastres optimisés avec l'inverse de CHIMERE par d'autres modèles pour simuler les mêmes épisodes de pollution est donc une étape permettant de déterminer le degré de dépendance au modèle de ces cadastres. Dans un premier temps, on peut envisager de tester les cadastres optimisés avec le même modèle mais avec des forçages différents, par exemple en remplaçant les données météorologiques de l'ECMWF par des champs issus de modèles tels que RAMS ou MM5.

### Quantifier l'incertitude des cadastres optimisés

Notre méthodologie d'inversion ne permet pas d'associer une incertitude aux cadastres optimisés obtenus. Le projet OPTEMI (OPTimisation des Emissions par Modélisation Inverse), dans lequel s'inscrit ce travail, prévoit des interactions entre l'inversion et des simulations Monte-Carlo. Ces dernières permettent en effet de quantifier les incertitudes des différents paramètres d'entrée du modèle.

### Application aux COV

Dans les cadastres actuels, les émissions de COV comportent une grande incertitude qui porte à la fois sur la quantité totale et sur la spéciation. Compte-tenu de l'importance du

rapport COV sur  $\text{NO}_x$  à l'émission dans la production d'ozone, l'optimisation indépendante des émissions de COV et de  $\text{NO}_x$  pourrait s'avérer très intéressante pour l'étude de la pollution photo-oxydante.

Des mesures de concentrations et surtout de spéciation de COV effectuées en routine par les réseaux seraient donc nécessaires. Bien que les normes imposent la prise en compte en routine d'espèces de plus en plus nombreuses en même temps que les mesures deviennent techniquement réalisables, on ne peut espérer disposer rapidement de données d'observation spécifiques aux COV assez nombreuses.

À défaut d'avoir accès à la spéciation, les mesures d'ozone devraient permettre d'inverser l'ensemble de la famille des COV. Bien qu'il ait été montré que les mesures d'ozone disponibles actuellement en Île-de-France n'étaient pas assez nombreuses pour contraindre efficacement l'inversion des émissions de COV, le développement du réseau de mesure de l'ozone dans les zones périphériques pourrait, dans des délais raisonnables, apporter les informations nécessaires.

### **Application à d'autres régions**

L'optimisation des émissions a un intérêt immédiat pour les régions qui disposent d'un réseau de mesure assez bien développé mais d'un cadastre peu précis ou non remis à jour. Cependant, même dans une région disposant, comme l'Île-de-France, d'un «bon» cadastre, l'optimisation des émissions permet de prendre en compte les jours particuliers, par exemple les grands départs en région parisienne. Les cadastres optimisés peuvent finalement être utilisés par les réseaux de surveillance de la qualité de l'air comme AIRPARIF, AIRFOBEP et AIRMARAIX, d'une part, dans le but d'obtenir de meilleures prévisions des pics et d'autre part, pour effectuer des simulations de scénarios d'émissions plus précis.

Le coût d'une optimisation fiable correspond à celui des stations de mesure nécessaires à l'obtention d'analyses de bonne qualité. Cette dernière constitue un enjeu dont le cadre dépasse largement celui de l'inversion. C'est pourquoi le coût matériel et financier de l'optimisation des émissions n'est peut-être pas systématiquement plus élevé que celui de la mise à jour d'un cadastre par des méthodes de type «bottom-up».

La méthode d'inversion a été utilisée en Île-de-France pour des raisons de simplicité de la dynamique et de disponibilité des observations. L'objectif est de l'appliquer à des sites de plus en plus complexes. Dans un premier temps, on étudie une région possédant un réseau de mesure dense mais dont le cadastre comporte des incertitudes plus grandes que celui de l'Île-de-France, par exemple en raison d'émissions importantes de COV biogéniques. On pourrait ensuite optimiser les émissions dans une région comportant également de fortes incertitudes sur d'autres paramètres, par exemple sur les champs dynamiques.

Actuellement, la région dans laquelle on dispose du plus grand nombre d'observations est celle de Fos-Berre Marseille où a eu lieu la campagne ESCOMPTE au cours des étés 2000 et 2001. Les cadastres d'émissions optimisés pourraient être utilisés par les équipes travaillant sur les analyses des épisodes de cette campagne. Malheureusement, la dynamique y est beaucoup plus complexe qu'en Île-de-France (phénomènes de brise de mer notamment). La durée exceptionnelle de la campagne de mesure permet néanmoins de disposer d'épisodes «simples» (régime de brise bien établi par exemple) qui peuvent constituer les premiers cas d'inversion.

## **Annexes**

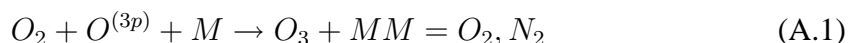
# Annexe A

## La photochimie dans la troposphère

Les réactions chimiques expliquant la production et la destruction de l’ozone dans la troposphère sont nombreuses. Cette annexe présente les principales étapes du cycle de l’ozone ( $O_3$ ) qui mettent en jeu le dioxyde d’azote ( $NO_2$ ), le monoxyde d’azote ( $NO$ ), deux espèces regroupées dans la famille des oxydes d’azote ( $NO_x$ ), le monoxyde de carbone ( $CO$ ), les composés organiques volatils (COV), les radicaux hydroxyles ( $OH$ ) et hydroperoxydes ( $HO_2$ ) au cours d’un processus rapide, l’**équilibre photochimique**, et d’un processus plus lent, le **cycle des radicaux**.

### A.1 L’équilibre photochimique

Une molécule d’ozone est formée par la combinaison d’un atome d’oxygène dans l’état fondamental ( $O^{(3p)}$ ) avec une molécule de dioxygène, un troisième corps permettant de stabiliser la réaction :



Dans la troposphère, l’atome  $O^{(3p)}$  provient le plus souvent d’une molécule de  $NO_2$  qui a été photodissociée par un rayonnement d’une longueur d’onde comprise entre 290 et 400 nanomètres (visible et proche ultra-violet) :



La molécule de  $NO$  formée peut elle aussi réagir avec une molécule d’ozone : il s’agit de la **titration** de l’ozone par  $NO$  qui donne :



Les équations A.1 et A.2 d’une part et A.3 d’autre part constituent l’**équilibre photochimique** et ne peuvent expliquer la production nette d’ozone qui conduit à des concentrations supérieures à 100 ppb lors de certains épisodes de pollution.

### A.2 Le cycle des radicaux

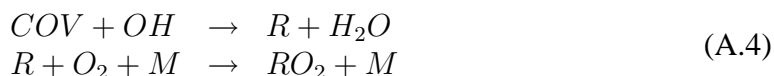
Une description détaillée du cycle décrit ici brièvement peut être consultée dans Toupance et al. (1988).

### A.2.1 Production d'ozone

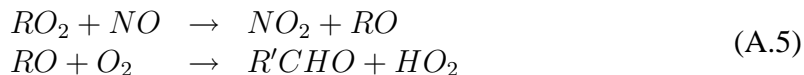
Lorsqu'un oxydant autre que l'ozone est disponible pour convertir le NO en NO<sub>2</sub> (réaction A.3), celui-ci peut s'accumuler dans la troposphère. Le cycle d'oxydation des COV ou **cycle des radicaux** convertit le NO en NO<sub>2</sub> ; ce dernier est alors photodissocié selon la réaction A.2 et l'ozone est produit suivant la réaction A.1.

Le cycle des radicaux n'est pas initié par les mêmes molécules en milieu urbain et en milieu rural. En milieu rural, les COV, le méthane et le CO peuvent être oxydés par OH en présence de NO et de rayonnement solaire. En milieu urbain, le méthane et le CO n'ont pas le temps de réagir avec OH, le temps de passage d'une masse d'air au-dessus d'une agglomération étant en général bien inférieur au temps de réaction de ces oxydations.

La première étape d'un cycle d'oxydation d'un COV consiste en la production d'un peroxy-radical (RO<sub>2</sub>) très réactif :



Ce peroxy-radical assure la conversion de NO en NO<sub>2</sub> en produisant un composé carbonyle R'CHO :



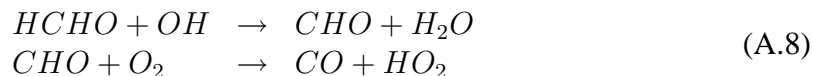
De plus, le radical HO<sub>2</sub> permet la conversion d'une autre molécule de NO en NO<sub>2</sub> et régénère un radical OH, disponible pour l'oxydation d'un autre COV :



Le bilan des réactions précédentes permet donc la conversion de deux molécules de NO en NO<sub>2</sub> :



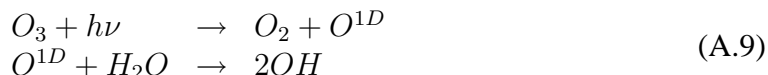
Le cycle n'est pas terminé puisque le composé R'CHO peut être oxydé par OH et produire un HO<sub>2</sub> disponible pour une seconde réaction A.6. L'exemple simple du formaldéhyde (HCHO) donne :



Les radicaux OH, RO<sub>2</sub> et HO<sub>2</sub> sont constamment recyclés. La production d'ozone est donc d'autant plus importante que le COV en jeu peut subir un grand nombre de cycles d'oxydation. Une molécule de méthane permet la formation de quatre molécules d'ozone mais une molécule de C<sub>5</sub>H<sub>12</sub> permet la formation de 12 molécules d'ozone. La production d'ozone dépend donc de la quantité de NO disponible et de la présence des radicaux.

### A.2.2 Sources et puits pour les radicaux

La principale source de production de radicaux OH est la photolyse de l'ozone par des rayonnements de longueur d'onde comprise entre 290 et 310 nanomètres :



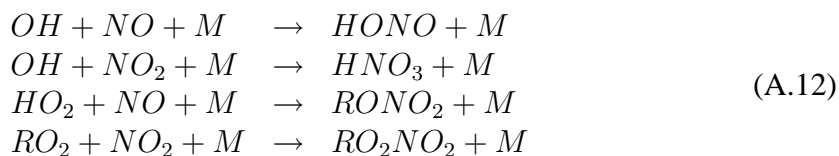
La photolyse de composés carbonylés est une source de radicaux  $HO_2$  ; par exemple avec le formaldéhyde, on obtient :



Deux puits de radicaux très différents existent. Les radicaux peuvent réagir entre eux :



Ils peuvent aussi réagir avec NO ou  $NO_2$ , notamment pour former de l'acide nitrique ( $HNO_3$ ) qui peut contribuer à l'acidification des précipitations et des nitrates organiques ( $RONO_2$ ) qui contribuent au transport du  $NO_2$  à longue distance :



### A.2.3 Les régimes chimiques

Dans un environnement pauvre en NO, par exemple en milieu rural, le puits de radicaux dominant est l'ensemble de réactions A.11. Une augmentation des émissions de  $NO_x$  permet donc d'augmenter proportionnellement la production d'ozone en fournissant des molécules de NO à la réaction A.5 avant que  $RO_2$  ne disparaissent avec d'autres radicaux. Dans un environnement riche en NO, par exemple en milieu urbain, le puits de radicaux dominant est l'ensemble des réactions A.12. Une augmentation des émissions de  $NO_x$  augmente donc la perte des radicaux et limite le cycle : la formation d'ozone est réduite, tant que les émissions de COV n'augmentent pas elles aussi, la réaction limitante étant A.4.

La production d'ozone est donc sensible au rapport  $COV/NO_x$  des émissions plutôt qu'à leur intensité. Différents **régimes chimiques** peuvent être observés selon le milieu. On parle de régime  $NO_x$  limité quand, pour des émissions de COV constantes, une augmentation des émissions de  $NO_x$  conduit à une augmentation de la production d'ozone. On parle de régime COV limité quand, pour des émissions de COV constantes, une augmentation des émissions de  $NO_x$  conduit à une diminution de la production d'ozone. Cette sensibilité particulière de la production d'ozone aux émissions de précurseurs explique que les pics de concentrations d'ozone sont observés non pas près des sources mais dans le panache des agglomérations. Pour une étude des différents régimes chimiques selon les concentrations de COV et de  $NO_x$ , voir Sillman (1999) ; pour une étude de la sensibilité des régimes chimiques, voir Menut et al. (2000).

# Annexe B

## La campagne ESQUIF

Toutes les informations concernant la campagne ESQUIF sont disponibles dans ESQUIF (2001) ainsi que dans plusieurs articles de synthèse dont Vautard et al. (2003, Menut et al. (2000). Les caractéristiques générales de la campagne et les situations rencontrées lors des Périodes d'Observations Intensives (POI) sont brièvement rappelées ici.

### B.1 La campagne de mesure

En plus des mesures de routine effectuées par le réseau AIRPARIF pour les concentrations de polluants et par METEO-FRANCE pour les paramètres météorologiques, un dispositif expérimental particulier a été mis en œuvre pour la campagne ESQUIF. Il comportait :

- des mesures aéroportées portant sur :
  - la structure tridimensionnelle des champs de polluants
  - les concentrations de fond
  - l'étude de la production photochimique d'ozone
  - l'étude des aérosols urbains.

Tous les détails des mesures aéroportées effectuées au cours des POI sont disponibles dans (*Vautard et al., 2001 ; Beekman et Derognat, 2003*).

- des mesures fixes au sol en un site urbain et deux sites péri-urbains. Celles-ci comprenaient :
  - des mesures de la dynamique atmosphérique (par LIDAR)
  - des mesures de concentrations de polluants (par spectrométrie DOAS)
  - des mesures de rayonnement et de taux de photolyse
  - des mesures du contenu en aérosols et de la granulométrie et de la composition chimique ceux-ci.
- des mesures mobiles au sol permettant de cartographier les champs de polluants et de déterminer les concentrations aux limites du domaine de certains COV.

## B.2 Caractéristiques des POI

Nous reprenons un tableau de synthèse (Tableau B.1), tiré de ESQUIF (2001), qui présente toutes les périodes de mesure de la campagne et donne quelques indications sur le type de situation rencontrée.

POI	Dates	Origine de la masse d'air	Remarques (sur toute la POI)
<u>Été 1998</u>			
1	25 → 26/07/1998	Continentale 1-2 jours de stagnation	Période nuageuse (moyens et hauts), vent faible
2	07 → 09/08/1998	Continentale 2-5 jours de stagnation	Ciel clair, fortes températures et vent faible
<u>Été 1999</u>			
3	16 → 17/06/1999	Continentale 3-4 jours de stagnation	Vent dominant de Nord-Est $< 3 \text{ m.s}^{-1}$
4	25 → 26/06/1999	Continentale 3-4 jours de stagnation	Ciel clair, vent venant souvent de l'Est $\approx 5 \text{ m.s}^{-1}$
5	01 → 02/07/1999	Maritime	Ciel clair, vent de Sud-Est $\approx 4 \text{ m.s}^{-1}$
6	16 → 18/07/1999	Maritime puis stagnant	Ciel clair
7	24 → 25/07/1999	Continentale 2-3 jours de stagnation	Expérience de suivi du panache parisien au-dessus de la Bretagne. Période ventée.
8	28 → 31/07/1999	Continentale 3-4 jours de stagnation	Vent très faible $< 2 \text{ m.s}^{-1}$
<u>Hiver 1999-2000</u>			
9	07 → 08/10/1999	Continentale 2-3 jours de stagnation	Conditions anticycloniques. Expériences pour la qualification de la conversion $\text{NO} \rightarrow \text{NO}_2$
10	26 → 28/01/2000	Continentale	Expérience pour l'observation de situations à très fortes $[\text{NO}_x]$ . Panache parisien vers l'Ouest.
<u>Été 2000</u>			
11	19 → 20/07/2000	Épisode photochimique modéré sans stagnation continentale	Expérience sur les aérosols
12	30 → 31/07/2000	Conditions sèches, pas de stagnation	Expérience sur les aérosols

TAB. B.1 – Récapitulatif des douze POI d'ESQUIF. La colonne «origine de la masse d'air» indique l'origine continentale ou maritime de celle-ci. Les «remarques» mettent en relief la spécificité de chacune des situations météorologiques.



# Annexe C

## L'adjoint

### C.1 Principe

Le principal objectif du modèle adjoint est le calcul du gradient  $\nabla H(X)$  de la fonction scalaire  $H(X)$ . Cette fonction, dans le cas de l'inversion, est la fonction coût soit une somme comprenant les concentrations chimiques de certaines espèces à certains instants et en certains lieux. Ce gradient dépend des variables pronostiques du modèle par rapport aux paramètres météorologiques (vent, température) et chimiques (émissions, taux de réactions).

On considère un ensemble de variables  $X$  évoluant spatialement et temporellement suivant le système d'équations suivant :

$$\frac{dX}{dt} = F(X, Y), \quad (\text{C.1})$$

$X$  représente donc les concentrations et est le «vecteur d'état» du système photochimique. Puisque le temps est discret dans le modèle, on peut noter ce vecteur d'état  $X_n$  à l'instant  $t = ndt$ . Le vecteur  $Y$  contient l'ensemble des paramètres. Enfin,  $F$  est la fonction (non linéaire) donnant la tendance instantanée de  $X_n$ , c'est-à-dire contenant toute la physico-chimie du modèle mais aussi ses approximations numériques. L'équation discrétisée d'évolution des variables du modèle s'écrit :

$$X_{n+1} = F(X_n, Y) \quad (\text{C.2})$$

Soit le gradient  $S$  de la fonction scalaire  $H(X)$  de  $X$  à un instant donné  $t$  :

$$S = \nabla_Y(H(X(t))). \quad (\text{C.3})$$

$S$  est aussi appelé sensibilité de  $H(X_n)$  par rapport à  $Y$ .

En se donnant un produit scalaire ( $\langle . \rangle$ ) dans l'espace des paramètres, le gradient  $S$  de  $H(X_n)$  par rapport à  $Y$  est tel que toute perturbation infinitésimale  $\delta Y$  de  $Y$  va créer une perturbation infinitésimale  $\delta H(X_n)$  de  $H(X_n)$ , telle que :

$$\delta H(X_n) = \langle S, \delta Y \rangle. \quad (\text{C.4})$$

À chaque instant  $t$ , la perturbation de  $X$  s'exprimera avec le système suivant :

$$\delta X_{n+1} = \left( \frac{\partial G}{\partial X}(X_n, Y) \delta X_n \right) + \left( \frac{\partial G}{\partial Y}(X_n, Y) \delta Y \right). \quad (\text{C.5})$$

En faisant l'hypothèse qu'il n'y a pas de perturbation initiale sur  $\delta X_0$ , on obtient, par remplacement successif dans (C.5), une expression pour la perturbation finale  $\delta X_n$  :

$$\delta X_n = A \delta Y, \quad (C.6)$$

où  $A$  est une matrice composée des sommes successives des itérations des dérivées partielles de la fonction  $G$ , par rapport à  $X$  et  $Y$ .

On a donc :

$$\delta H(X_n) = \frac{\partial H}{\partial X} A \delta Y, \quad (C.7)$$

ce qui peut, aussi s'écrire, en utilisant l'opérateur d'adjonction (noté par une  $*$ ) :

$$\delta H(X_n) = \left\langle A^* \frac{\partial H}{\partial X}, \delta Y \right\rangle. \quad (C.8)$$

Dans ce cas, on notera que l'opérateur d'adjonction est identique à une transposition car le produit scalaire choisi est canonique. Cette dernière relation, eq. (C.8) montre en particulier (si on la compare à l'équation (C.4)) que :

$$S = A^* \left( \frac{\partial H}{\partial X} \right)^* \quad (C.9)$$

L'intégration «à rebours» dans le temps des équations adjointes du système (C.2) peut résoudre numériquement le système (C.9) tel que, avec

$$\delta^* X_n = \left( \frac{\partial H}{\partial X} (X_n) \right)^*$$

et

$$\delta^* X_{n-1} = \left( \frac{\partial G}{\partial X} (X_{n-1}, Y) \right)^* \delta^* X_n; \quad (C.10)$$

la valeur finale du calcul  $\delta^* X_0$  représentera le gradient (ou la sensibilité) de  $H$  par rapport aux conditions initiales. Le gradient  $S$  par rapport aux paramètres sera alors donné par la condition finale du système, c'est-à-dire :

$$\begin{cases} \delta^* Y_n &= 0 \\ \delta^* Y_{n-1} &= \delta^* Y_n + \left( \frac{\partial G}{\partial Y} (X_{n-1}, Y) \right)^* \delta^* X_n \\ S &= \delta^* Y_0 \end{cases} \quad (C.11)$$

En pratique, les modèle adjoints relatifs aux systèmes C.10 et C.11 sont intégrés simultanément, la seule différence étant que le gradient est cumulé au sein d'un pas de temps physique.

## C.2 Programmation

### C.2.1 Principe d'écriture d'un code adjoint

Le modèle adjoint de CHIMERE a été développé «ligne à ligne» de code. Pour chaque «instruction» représentant un processus physique, l'équation est différenciée au premier ordre, linéarisée. On obtient donc le modèle «linéaire tangent». Le passage du code «linéaire tangent» au code «adjoint» reste dans notre cas une transposition des opérations. Pour fixer les idées, la procédure est décrite sur un exemple simple ci-dessous.

#### □ Code direct

$$result = val1 + val2 \times val3 + val4$$

avec  $val1$ ,  $val2$  et  $val3$  des paramètres du modèle que l'on souhaite «adjointiser» et  $val4$  une constante.

On linéarise autour des valeurs de  $val1$ ,  $val2$  et  $val3$  pour obtenir le linéaire tangent.  $val4$  étant une constante, sa dérivée est nulle et n'apparaîtra donc pas dans la suite du calcul. Les variables  $val1$ ,  $val2$  et  $val3$  sont donc des **variables directes d'entrée** et  $result$  est le résultat, i.e. la **variable directe de sortie** de la procédure de calcul.

#### □ Code linéaire tangent

$$dresult = dval1 + val2 \times dval3 + dval2 \times val3$$

Pour finalement passer du code linéaire tangent au code adjoint, on utilise la propriété qu'une ligne de l'opérateur linéaire direct devient par transposition (hypothèse précédente sur le produit scalaire canonique), une colonne de l'opérateur adjoint. On prend donc le coefficient devant chacune des variables d'entrée (i.e.  $dval1$ ,  $dval2$  et  $dval3$ ) pour le mettre en facteur de la variable de sortie du linéaire tangent (i.e.  $dresult$ ).

#### □ Code adjoint

$$\begin{cases} aval3 &= aval3 + aval2 \times aresult \\ aval2 &= aval2 + aval3 \times aresult \\ aval1 &= aval1 + aresult \end{cases}$$

Dans ce cas, on a  $areult$  la **variable adjointe d'entrée** et  $aval1$ ,  $aval2$  et  $aval3$  les **variables adjointes de sortie**. Pour respecter l'ordre d'actualisation des variables (extrêmement important puisque l'on cumule les valeurs des variables adjointes), on notera que l'ordre des relations précédentes a été inversé par rapport à la formulation directe. Cela est particulièrement important lorsque, au sein de la même procédure, une variable voit sa valeur réaffectée plusieurs fois.

La validation de ces codes linéaire tangent et adjoint peut se faire de manière automatique, en utilisant les propriétés des résidus d'écart. Le premier d'entre eux se calcule entre le code direct et le code linéaire tangent. Pour une perturbation donnée  $\delta x$  de la variable d'état, ce résidu est calculé comme l'écart entre la valeur du code direct au point de perturbation et la valeur du code direct au point sans perturbation augmenté de l'approximation par le code linéaire tangent. Symboliquement, avec des notations naturelles :

$$Residu = H(x + \delta x) - H(x) - \langle LTH(x), \delta x \rangle$$

Ce résidu doit évidemment être du premier ordre par rapport aux variations des perturbations `dval1`, `dval2`, `dval3` et `dval4`. Le second résidu se calcule entre le code linéaire tangent et le code adjoint, qui placés à la suite dans un ordre ou dans l'autre, donnent le même résultat, à la précision machine utilisée. Cela exprime le fait que les deux opérateurs sont adjoints. Le troisième résidu provient de la même comparaison que dans le second, mais cette fois entre le code direct et le code adjoint. Le résidu sur le calcul de la valeur de fonctionnelle  $J$  doit alors être du premier ordre aussi par rapport aux perturbations.



### C.2.3 Vérification informatique du code adjoint

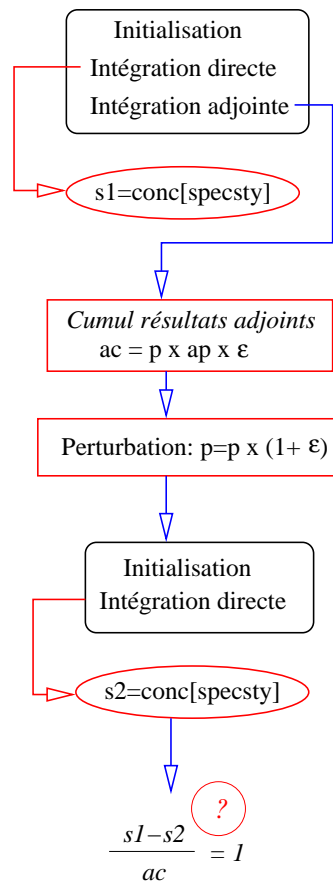


FIG. C.2 – Schéma de principe de la vérification du codage du modèle adjoint

Une méthode simple est mise en place pour vérifier la validité du code adjoint (Figure C.2). Cette méthode s'apparente à un développement de Taylor au premier ordre. Au premier ordre car, de toute façon, l'adjoint est construit à partir du linéaire tangent, qui est lui-même une dérivée du premier ordre seulement.

## C.3 Schéma de transport de Van Leer : codes direct et adjoint

Le schéma de transport de Van Leer permet de calculer le transport d'une espèce chimique d'une boîte du maillage à une autre dans un plan horizontal. Il est ici codé en FORTRAN.

### C.3.1 Code direct

```

function vanleer(cupup,cup,cdo,dx,u,dt)

implicit none
double precision cupup,cup,cdo,u,dt,cfl,vanleer,minmod,dx

cfl = u*dt/dx
if (cfl.gt.1.) then
  print *, '*** CFL > 1 IN VANLEER scheme'
  print *, ' u (m/s) dt (s) dx (km) cfl'
  print *, ' ',u/100.,dt,dx/1.e6,' ',cfl
  if(cfl.ge.2)stop
endif
c Nonmonotonic cases
if(cup.lt.min(cupup,cdo)) then
  vanleer = cup
elseif(cup.gt.max(cupup,cdo)) then
  vanleer = cup
else
c Monotonic cases
  if(abs(cdo-cup).lt.abs(cup-cupup)) then
    minmod = cdo - cup
  else
    minmod = cup - cupup
  endif
  vanleer = cup + 0.5*(1. - cfl)*minmod
endif

return
end

```

### C.3.2 Adjointisation

La variable directe *vanleer* est une concentration dépendant des variables d'entrée directes *cupup* (var. eq. 1), *cup* (var. eq. 2) et *cdo* (var. eq. 3). On cherche donc à obtenir les variables de sortie adjointes  $\delta cupup^*$ ,  $\delta cup^*$  et  $\delta cdo^*$  en fonction de la variable adjointe d'entrée  $\delta vanleer^*$ .

Traitons le cas monotone dans lequel on a  $|cdo - cup| < |cup - cupup|$ . On obtient alors pour la variable directe *vanleer* :

$$vanleer = cup + 0.5 \times (1. - cfl) \times (cdo - cup), \text{ soit :}$$

$$vanleer = cup \times -0.5 \times (1. + cfl) + 0.5 \times (1. - cfl) \times cdo$$

Le code linéaire tangent

correspondant, obtenu en linéarisant autour de chacune des trois variables, est donc :

$$\delta vanleer = \delta vanleer_{cupup} \times \delta cupcup + \delta vanleer_{cup} \times \delta cup + \delta vanleer_{cdo} \times \delta cdo \text{ avec :}$$

$$- \delta vanleer_{cupup} = 0$$

$$- \delta vanleer_{cup} = -0.5 \times (1. + cfl)$$

$$- \delta vanleer_{cdo} = 0.5 \times (1. - cfl)$$

On obtient donc le code adjoint suivant :

$$- \delta cupup^* = \delta vanleer_{cupup} \times \delta vanleer^*$$

$$- \delta cup^* = \delta vanleer_{cup} \times \delta vanleer^*$$

$$- \delta cdo^* = \delta vanleer_{cdo} \times \delta vanleer^*$$



### C.3.3 Code adjoint

```

function dvanleer(var,cupup,cup,cdo,dx,u,dt)

implicit none
double precision cupup,cup,cdo,u,dt,cfl,dvanleer,minmod,dx
integer var

cfl = u*dt/dx
if (cfl.gt.1.) then
  print *, '*** CFL > 1 IN VANLEER scheme'
  print*, ' u (m/s) dt (s) dx (km) cfl'
  print*, ' ',u/100.,dt,dx/1.e6,' ',cfl
  if(cfl.ge.2)stop
endif

c Nonmonotonic cases
if(cup.lt.min(cupup,cdo)) then
  if(var.eq.2) then
    dvanleer=1.
  else
    dvanleer=0.
  endif
elseif(cup.gt.max(cupup,cdo)) then
  if(var.eq.2) then
    dvanleer=1.
  else
    dvanleer=0.
  endif
else
  c Monotonic cases
  if(abs(cdo-cup).lt.abs(cup-cupup)) then
    if(var.eq.1) then
      dvanleer=0.
    elseif(var.eq.2) then
      dvanleer=1-0.5*(1.-cfl)
    elseif(var.eq.3) then
      dvanleer=0.5*(1.-cfl)
    endif
  else
    if(var.eq.1) then
      dvanleer=-0.5*(1.-cfl)
    elseif(var.eq.2) then
      dvanleer=1+0.5*(1.-cfl)
    elseif(var.eq.3) then
      dvanleer=0.
    endif
  endif
endif

return
end

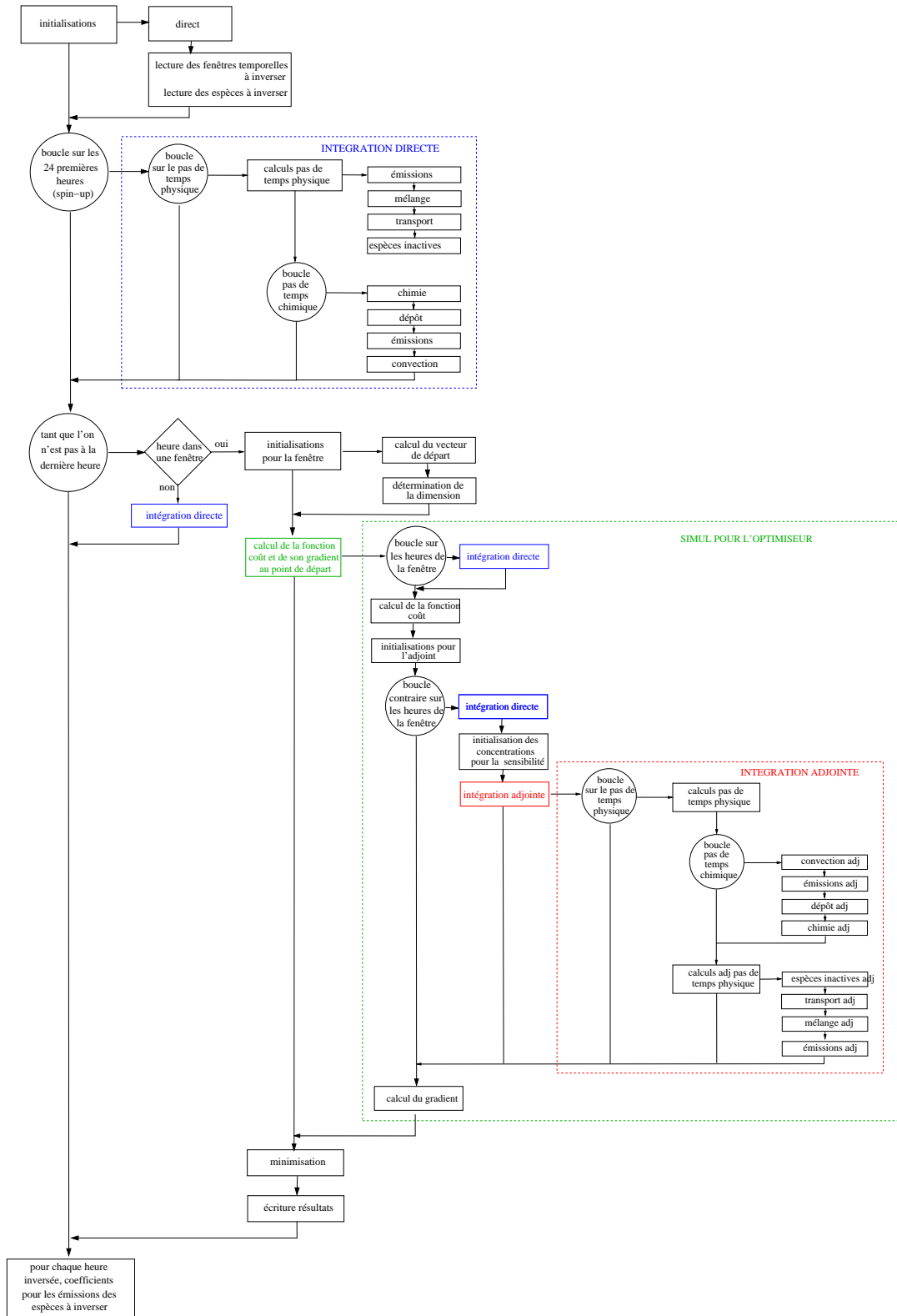
```

# **Annexe D**

## **Inverse**

Le modèle CHIMERE V200211K direct compte environ 3500 lignes de programme ; les routines de l'adjoint même correspondent à environ 2000 lignes ; l'optimiseur N1QN3 comporte un peu plus de 1300 lignes de code ; enfin, les routines propres à l'inverse (qui font appel à toutes les autres) représentent environ 4200 lignes de code.

L'algorithme de l'inverse est représenté sur la Figure D.1. Le krigeage et l'agrégation spatiale, qui sont effectués en amont de l'inversion, n'y figurent pas.

FIG. D.1 – *Algorithme de l'inverse de CHIMERE.*

# Annexe E

## Le krigeage

*Le matériel et les notations de cette annexe sont tirés de Blond (2002a) pp. 38 à 52.*

### E.1 Le krigeage

#### E.1.1 Principe

Le krigeage permet d'obtenir en chaque point de l'espace  $s_a$  la concentration estimée  $Z^a(s_a)$  grâce à une combinaison linéaire de  $p$  observations  $Y^0$  effectuées aux points  $s_k, k = 1, p$  :

$$Z^a(s_a) = \sum_{k=1}^p w_k(s_a) Y^0(s_k) \quad (\text{E.1})$$

où  $w_k(s_a)$  est le poids de l'observation  $Y^0(s_k)$  dans l'estimation de la concentration vraie  $Z^t(s_a)$ .

#### E.1.2 Les poids

On suppose que l'on cherche à estimer un champ de concentration uniforme en moyenne sur le domaine considéré mais prenant des valeurs aléatoires en chaque point. On a donc  $E(Z^t) = m$  où  $E()$  indique la moyenne et  $m$  est une constante. On suppose de plus que les observations ne sont pas biaisées. On a donc :  $E(Y^0(s_k)) = E(Z^t(s_k)) = m$ .

La moyenne de l'erreur d'estimation  $\epsilon_a(s_a)$  s'exprime alors selon :

$$\begin{aligned} E[\epsilon_a(s_a)] &= E[Z^a(s_a)] - E[Z^t(s_a)] \\ &= \sum_{k=1}^p w_k(s_a) E[Y^0(s_k)] - E[Z^t(s_a)] \\ &= m \left( \sum_{k=1}^p w_k(s_a) - 1 \right) \end{aligned}$$

Pour obtenir une estimation  $Z^a(s_a)$  de  $Z^t(s_a)$  non biaisée, il faut donc que les poids satisfassent la condition :

$$\sum_{k=1}^p w_k(s_a) = 1 \quad (\text{E.2})$$

Si cette condition est remplie,  $Z^a(s_a)$  est la meilleure estimation possible de  $Z^t(s_a)$  si la variance de son erreur  $\epsilon_a(s_a)$ , notée  $\sigma_a^2(s_a)$ , est minimale.

On a :

$$\sigma_a^2(s_a) = E\left[\sum_{k=1}^p w_k(s_a)(Y^0(s_k) - Z^t(s_a))\right]^2 \quad (\text{E.3})$$

$$= \sum_{k=1}^p \sum_{l=1}^p w_k(s_a)w_l(s_a)E[(Y^0(s_k) - Z^t(s_a))(Y^0(s_l) - Z^t(s_a))] \quad (\text{E.4})$$

En notant :

- $C_{k,l}$  les covariances du champ réel entre deux points  $s_l$  et  $s_k$
- $R_{k,l} = R_{k,l}\delta_{k,l}$  les covariances entre les erreurs d'observations, non corrélées entre elles pour des observations effectuées avec des instruments différents (réseau de mesure)

l'équation E.4 devient :

$$\sigma_a^2(s_a) = C_{a,a} - 2 \sum_{k=1}^p w_k(s_a)C_{a,k} + \sum_{k=1}^p \sum_{l=1}^p w_k(s_a)w_l(s_a)(C_{k,l} + R_{k,l}\delta_{k,l}). \quad (\text{E.5})$$

Pour obtenir une variance minimale, il faut minimiser la fonction  $J(s_a)$  donnée par :

$$J(s_a) = \sigma_a^2(s_a) + 2\mu\left(\sum_{k=1}^p w_k(s_a) - 1\right)$$

avec  $\mu$  multiplicateur de Lagrange. La minimisation de  $J(s_a)$  par rapport aux poids et au paramètre  $\mu$  conduit donc finalement au système :

$$\forall k = 1, \dots, p \quad \sum_{l=1}^p w_l(s_a)(C_{k,l} + R_{k,l}\delta_{k,l}) + \mu = C_{k,a}. \quad (\text{E.6})$$

Pour résoudre le système E.6, il est donc nécessaire de connaître les covariances  $C_{k,l}$ . Comme cela n'est pas possible, elles devront être modélisées.

Enfin, si l'on dispose des poids optimaux, on obtient la variance de l'erreur d'analyse en remplaçant les solutions de E.6 dans E.5 :

$$\sigma_a^2(s_a) = C_{a,a} - \sum_{k=1}^p w_k(s_a)C_{a,k} - \mu \quad (\text{E.7})$$

## E.2 Le krigeage des innovations

### E.2.1 Principe

L'interpolation statistique permet d'obtenir en chaque point de l'espace  $s_a$  la concentration analysée  $Z^a(s_a)$  en ajoutant à la concentration simulée  $Z^b(s_a)$  une combinaison linéaire des différences entre les  $p$  observations  $Y^0$  effectuées aux points  $s_k, k = 1, p$  et les concentrations simulées interpolées en ces points  $Z^b(s_k)$ . On a alors :

$$Z^a(s_a) = Z^b(s_a) + \sum_{k=1}^p w_k(s_a)(Y^0(s_k) - Z^b(s_k)) \quad (\text{E.8})$$

où  $w_k(s_a)$  est le poids de l'observation  $Y^0(s_k)$  dans l'estimation de la concentration vraie  $Z^t(s_a)$ .

L'équation de l'interpolation statistique E.8 peut être vue comme une équation de krigeage de la forme E.1 estimant le champ d'erreur du modèle  $Z^t - Z^b$  grâce aux «observations»  $X^0(s_k) = Y^0(s_k) - Z^b(s_k)$ . Les  $X^0(s_k)$  sont appelées innovations et l'analyse  $X^a(s_a) = Z^a(s_a) - Z^b(s_a)$  s'écrit alors :

$$X^a(s_a) = \sum_{k=1}^p w_k(s_a) X^0(s_k). \quad (\text{E.9})$$

### E.2.2 Les poids

Si l'on admet que les erreurs d'observation et les erreurs du modèle sont décorréliées (observations et simulations sont des estimations indépendantes), les poids du krigeage des innovations sont solutions du système :

$$\forall k = 1, \dots, p \quad \sum_{l=1}^p w_l(s_a) (B_{k,l} + R_{k,l} \delta_{k,l}) = B_{k,a}. \quad (\text{E.10})$$

où les  $B_{k,l}$  sont les covariances d'erreur du modèle. On a également :

$$\sigma_a^2(s_a) = B_{a,a} - \sum_{k=1}^p w_k(s_a) B_{a,k} - \mu \quad (\text{E.11})$$

## E.3 Krigeage intrinsèque

Cette technique suppose que la variable à analyser  $X^t$  est intrinsèquement stationnaire c'est-à-dire que :

$$\begin{aligned} \forall (s_k, s_l), E(X^t(s_1) - X^t(s_2)) &= 0 \\ \forall (s_k, s_l), \frac{1}{2} \text{var}(X^t(s_1) - X^t(s_2)) &= \gamma(s_1 - s_2) \end{aligned}$$

où  $\gamma(s_1 - s_2)$  est la semi-variance des différences entre  $X^t(s_1)$  et  $X^t(s_2)$ .

On utilise  $\gamma_{k,l} = \frac{1}{2} \text{var}(X^t(s_k) - X^t(s_l))$  pour modéliser  $B_{k,l}$  (E.10). Pour cela, on construit un variogramme des  $\gamma_{k,l}$ . Pour une fonction de covariance  $\gamma$  isotrope, pour  $J$  classes de distance  $N_j$  définies par l'ensemble des couples de stations éloignées de la distance  $h_j$ , on a :

$$\begin{aligned} \gamma^o(h_j) &= \frac{1}{2} \frac{1}{|N_j|} \sum_{N_j} (X^0(s_k) - X^0(s_l))^2 \quad \forall j = 1, \dots, J \\ &= \gamma_{k,l} + \frac{1}{4} (R_{k,k} + R_{l,l} - 2R_{l,k} \delta_{lk}) \end{aligned}$$

Si l'on suppose que les variances d'erreur d'observation sont homogènes et égales à  $\sigma_0^2$ , on obtient d'une part, à distance nulle,  $\gamma^o(0) = \gamma_{k,l} = 0$  et d'autre part, à distance non nulle,  $\gamma^o(0) = \gamma_{k,l} + \sigma_0^2$ . Le variogramme expérimental donne une estimation de  $\gamma_{k,l}$  à  $\sigma_0^2$  près. Il est donc modélisé par une fonction de la variance de l'erreur d'observation, ici :  $\gamma(h) = c(1 - \exp(-\frac{h}{a})) + \sigma_0^2$  avec  $a$  et  $c$  deux paramètres.

# Annexe F

## Méthode itérative et optimalité

### F.1 Définitions générales

- $M$  modèle, supposé parfait et inversible de taille  $m \times m$
- $Y$  vecteur des concentrations de taille  $m$
- $Y_0$  vecteur des observations de taille  $mo$
- $R_0$  matrice contenant les variances des analyses diagonale de taille  $mo \times mo$
- $X$  vecteur des émissions de taille  $m$   
 $\rightarrow Y = MX$  et  $X = M^{-1}Y$
- $J$  fonction coût :  $J(X)$
- $H$  opérateur permettant de ramener les concentrations simulées aux points d'observation (poids prenant en compte les quatre valeurs simulées autour de l'observation :  $h_{ij} = 0 \forall j \notin \{1, 2, 3, 4\} h_{ij} \neq 0 \forall j \in \{1, 2, 3, 4\}$ ) taille  $mo \times m$
- $W$  matrice des poids du krigeage de taille  $m \times mo$
- $Y_a$  vecteur des analyses de taille  $m \times m$  tel que :  $Y_a = Y + W(Y_0 - HY)$
- $R_a$  matrice contenant les variances des analyses diagonale de taille  $m \times m$

On suppose que le problème est bien posé c'est-à-dire qu'on dispose d'autant d'observations que de flux à inverser, soit  $m = mo$ . Pour simplifier, on considère de plus que les observations et les points de grille coïncident. Ainsi a-t-on  $H = I$ .

### F.2 Approche classique

La fonction coût  $J(X)$  est de la forme :  $J(X) = (Y(X) - Y_0)^T R_0 (Y(X) - Y_0)$ .

La solution exacte  $\hat{X}$  à  $J(X) = 0$  est donnée par :  $\hat{X} = M^{-1}Y_0$ .

### F.3 Développement : méthode itérative

#### F.3.1 Krigeage des concentrations

Soient  $Y^{fg}$  les concentrations simulées *a priori* avec les émissions *a priori*  $X^{fg}$ . Le krigeage donne le champ analysé  $Y_a^{fg}$  tel que :  $Y_a^{fg} = Y^{fg} + W(Y_0 - Y^{fg})$ .

### F.3.2 Inversion des émissions

La fonction coût  $J(X)$  est de la forme :  $J(X) = (Y(X) - Y_a)^T R_a (Y(X) - Y_a)$ .

Ici, on cherche les émissions optimisées  $X_{opt}^1$  telles que

$J(X_{opt}^1) = (Y(X_{opt}^1) - Y_a^{fg})^T R_a^{fg} (Y(X_{opt}^1) - Y_a^{fg})$  soit minimale. Le modèle  $M$  étant inversible, on peut écrire la solution exacte  $X_{opt}^1$  telle que  $J(X_{opt}^1) = 0$ . On a alors :

$$Y(X_{opt}^1) = Y_a^{fg} \text{ soit } X_{opt}^1 = M^{-1} Y_a^{fg}.$$

### F.3.3 Deuxième cycle krigeage-optimisation

Le krigeage est effectué avec les observations et les concentrations  $Y^1$  simulées avec les émissions  $X_{opt}^1$  soit :  $Y_a^1 = Y^1 + W(Y_0 - Y^1) = M X_{opt}^1 + W(Y_0 - M X_{opt}^1)$  ou encore, en fonction des résultats du précédent krigeage :  $Y_a^1 = Y_a^{fg} + W(Y_0 - Y_a^{fg})$ .

Les émissions optimisées  $X_{opt}^2$  sont données par :  $Y(X_{opt}^2) = Y_a^1$  soit  $X_{opt}^2 = M^{-1} Y_a^1$ .

### F.3.4 N-ième cycle krigeage-optimisation

On a, pour les concentrations analysées :  $Y_a^n = Y^n + W(Y_0 - Y^n)$  ou encore, en fonction des résultats du précédent krigeage,

$$Y_a^n = Y_a^{n-1} + W(Y_0 - Y_a^{n-1})$$

et pour les émissions :

$$X_{opt}^{n+1} = M^{-1} Y_a^n.$$

### F.3.5 Convergence

On a :

$$Y_a^n = W Y_0 + (I - W) Y_a^{n-1}$$

On note  $G = I - W$  soit  $g_{i,j} = \delta_{ij} - w_{i,j}$ . En posant  $Z_n = Y_a^n - Y_0$ , on obtient :

$$Z_n = (I - W) Z_{n-1} = (I - W)^n Z_{fg} = G^n Z_{fg}.$$

La convergence de la méthode est donc liée à celle de la suite de matrices  $(G^n)$ .



# Annexe G



AE International – Europe

Atmospheric Environment 38 (2004) 971–983

**ATMOSPHERIC  
ENVIRONMENT**

[www.elsevier.com/locate/atmosenv](http://www.elsevier.com/locate/atmosenv)

## Quantification of the impact of aircraft traffic emissions on tropospheric ozone over Paris area

Isabelle Pison\*, Laurent Menut

*Laboratoire Inter-Universitaire des Systèmes Atmosphériques, 61, av. du Général de Gaulle, UMR CNRS 7583, Universités Paris 7, Paris 12, 94010 Créteil Cedex, France*

Received 3 June 2003; received in revised form 27 October 2003; accepted 31 October 2003

### Abstract

Accurate estimations of the emissions of primary pollutants are crucial for the modeling of photo-oxidants' concentrations. For a majority of chemistry-transport models (CTMs), these emissions are taken into account near the surface only. They are expressed as surface fluxes and represent surface activities such as traffic, industries or biogenic processes. However, in the vicinity of large cities, commercial aircraft emissions represent a nonnegligible source, located both at the surface and at altitude, including landing and take-off of aircraft within the boundary layer. This is the case of Paris where one national airport (Le Bourget) and two international airports (Roissy-Charles-de-Gaulle and Orly) are located less than 30 km away from the city center. This study presents the first-model analysis of the impact of aircraft emissions on photo-oxidant concentrations over the Paris area. Using a three-dimensional aircraft emission inventory, we compare ozone surface concentrations obtained with and without these emissions by running the CTM CHIMERE during the second Intensive Observation Period of the ESQUIF project. The simulated differences enable us to estimate the impact of aircraft traffic emissions on ozone surface concentrations in and around the city. The results showed that the maximum impact, which consists in a fast ozone titration by NO near the airports within the surface layer, occurs during the night. In remote areas and at altitude, adding new emissions enhanced photo-chemistry during the afternoon. In order to estimate the impact of the uncertainty of our inventory, aircraft emitted masses of volatile organic compounds (VOCs) and NO<sub>x</sub> are perturbed. The results showed that NO<sub>x</sub> air traffic emissions have a more important impact than VOC emissions, particularly during the night and near the sources. Nevertheless, these variations of air traffic emissions do not change previous conclusions.

© 2003 Elsevier Ltd. All rights reserved.

**Keywords:** Urban air pollution; Three-dimensional modeling; Regional chemistry-transport model; Ozone sensitivity; Emission inventory

### 1. Introduction

Within the low troposphere, high ozone concentrations are a problem for public health. Ozone is formed by complex photo-chemistry involving primary pollutants such as nitrogen oxides (NO<sub>x</sub>) and volatile organic compounds (VOCs). These primary pollutants are directly emitted at the surface by anthropogenic

activities (traffic, solvents, industries and other activity sectors) and biogenic sources (such as cultivated fields and forests).

To understand the behavior of this photo-chemistry, chemistry-transport models (CTMs) are now widely used. Most of them are able to reproduce surface ozone concentration fields, which are then used for the analysis of high pollution events, scenarios on surface emissions and forecast. But these ozone concentration fields are not always as accurate as it is needed to help to air policy elaboration. This is due to (i) the fact that ozone

\*Corresponding author. Fax: +33-1-4517-1564.

E-mail address: [pison@lisa.univ-paris12.fr](mailto:pison@lisa.univ-paris12.fr) (I. Pison).

concentrations correspond to a budget involving numerous factors such as emissions, meteorology and (ii) the nonlinear interactions of all these factors.

Among all these factors, it was previously shown that the emissions of primary pollutants over urban areas are the most important as well as the most uncertain. For most CTM, these emissions are taken into account near the surface only. They are expressed as surface fluxes due to surface activities such as traffic, industries or biogenic processes. However, surface emissions are not the only primary source in the neighborhood of large cities: commercial aircraft emissions represent a non-negligible source, located both at the surface and at altitude, including landing and take-off of aircraft within the boundary layer. Moreover, the activity of international airports increases every year, which leads to a continuous increase of the emissions due to aircraft.

Numerous studies have been made to estimate the impact of air traffic emissions in the troposphere but only within its upper part (Gardner et al., 1997; Meijer et al., 2000; Colvile et al., 2001). These studies showed, for example, that about 60% of aircraft  $\text{NO}_x$  are emitted between 10 and 12 km at altitude, 93% of which in the Northern Hemisphere. It was also shown that aircraft emissions may represent up to 60% of  $\text{NO}_x$  emissions in the upper troposphere and could lead to an increase of a few percent of ozone global concentrations in this atmospheric layer (Derwent, 1982).

Studies concerning the lower troposphere, including the atmospheric boundary layer (ABL), are rare. Colvile et al. (2001) showed that the impact of aircraft emissions is poorly known but is mainly due to aircraft movements at the ground and related activities (buses on airports, taxi ways, etc.) rather than to the flights themselves. Moussiopoulos et al. (1997) showed that  $\text{NO}_x$  emissions are the most important contribution of aircraft emissions but high concentrations are only observed very close to the sources, i.e. to the airports. Finally, Perl et al. (1997) showed that the critical point for aircraft emissions is not the landing and take-off phases but all the stages during which aircraft wait at the ground (the “taxi” or “idle” phase). During this phase, engines do not work at their optimum conditions thus emitting significant amounts of carbon monoxide and hydrocarbons. However, none of these previous papers precisely quantified the impact of aircraft emissions on photo-oxidant concentrations within the ABL.

As a complement to these studies, we propose in this paper a verification and a direct quantification of the impact of aircraft traffic emissions in the ABL on ozone concentrations. This quantification is done using a new emission inventory specifically dedicated to air traffic around the large urban area of Paris. In this area, three major commercial airports are located less than 30 km from the city center, one national (Le Bourget) and two international airports (Roissy-Charles-de-Gaulle,

hereafter called CdG, and Orly). The methodology may nonetheless apply to any large city with suburban airports. The air traffic emission inventory is added to surface emissions fluxes usually taken into account in a regional CTM. For a real fully documented pollution event, we evaluate how much aircraft emissions during landing, take-off and taxi affect ozone concentrations in the Paris area.

## 2. Methodology

The Paris area is an urban area surrounded by three airports including two international ones. To study the impact of these airports, a real pollution event was selected; this corresponds to the second Intensive Observation Period (IOP2) of the ESQUIF campaign (Menut et al., 2000). To add new emissions in CHIMERE, a new air traffic emission inventory was elaborated. To understand all these steps, we first present the CHIMERE model, the usual surface emissions used, the studied case to finally explain how the new inventory was built.

### 2.1. The chemistry-transport model: CHIMERE

#### 2.1.1. General description of the model

CHIMERE is a CTM designed for the simulation of photo-oxidant concentrations within the ABL. Its applications concern the analysis and the real-time forecast of pollution events (the PIONEER project) as well as ozone climatology (Schmidt et al., 2001; Vautard et al., 2001). The model is developed over two domains, a macroscale one and a mesoscale one, and it is also suitable for pollution event analysis (Schmidt and Martin, 2003; Beekmann and Derognat, 2003).

For this study, the mesoscale version is used and chemical boundary conditions are ensured by one-way nesting within the continental version (see Vautard et al. (2001) and Schmidt and Martin (2003) for a complete description of the model and the whole set of parameterizations). Horizontally, the  $150 \times 150$  km domain is centered on Paris (see Fig. 1) and the model mesh is constituted by  $25 \times 25$  cells. Vertically, the mesh is composed by five levels, thicknesses of which increase with increasing altitude (0–50–650–1280–2030–3100 m) so as to be as accurate as possible within the surface layer. Despite this limited number of levels, the model was proved relevant over long periods, including the studied case (Vautard et al., 2001).

#### 2.1.2. Chemistry and surface emissions

CHIMERE uses MELCHIOR chemical mechanism, which counts 44 model species reacting through 116 chemical reactions (Lattuati, 1997). Emissions are taken into account as surface fluxes that are available hourly in

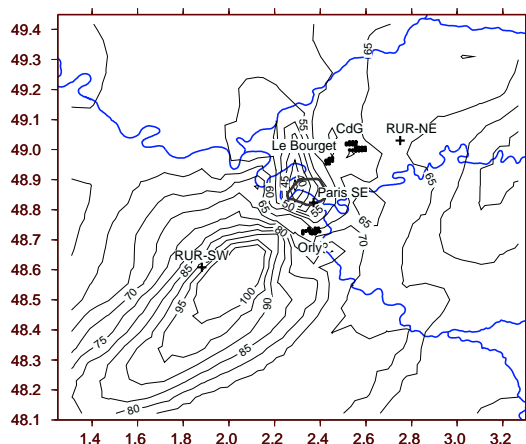


Fig. 1. Ozone concentrations (ppb) at the first level [0–50] m on 7 August 1998 at 3 p.m. The wind was very regular, coming in the domain from northeast and going out to northeast–east. Are also shown the locations of the three airports (CdG, Orly and Le Bourget), two selected points in rural areas, northeast (RUR-NE) and southwest (RUR-SW) from the town, and a point in the urban area, Paris southeast.

each grid cell. They are based on two data sources: (i) the CITEPA 1993 inventory, which provides annual total emissions of  $\text{NO}_x$ ,  $\text{SO}_2$ ,  $\text{CO}$ ,  $\text{CH}_4$  and nonmethane VOCs (not speciated), and (ii) the 1994 Generation for European Emission Data for Episodes (GENEMIS) project database, which provides VOC speciation and time distribution of annual total masses into hourly values. Note that these surface inventories already take surface emissions due to airport activity (e.g. road traffic) into account but do not include emissions due to aircraft. The two steps of the construction of model emission inventory are briefly described here. Supplementary information can be found in Vautard et al. (2001).

The first step consists of speciation and temporal de-aggregation of annual total emissions. The GENEMIS VOC speciation is used to obtain a speciation in 32 NAPAP classes (Middleton et al., 1990) for each activity. Temporal variations are calculated on the basis of GENEMIS data for three typical days (weekdays, Saturday and Sunday), for each calendar month, each hour and each activity sector. Only the weekday diurnal profile for traffic is not taken from the GENEMIS database but from a recent publication of the emission profile of  $\text{CO}$  in the Paris area (Sallès et al., 1996).

Once the de-aggregation is performed, an aggregation step to lump VOCs into model species is achieved (Middleton et al., 1990). A time-integrated OH concentration (INTOH) of 1.0 ppt.min is used as a control parameter value. Simulations were done to test this

value of INTOH and concentrations appeared quite insensitive to INTOH changes. Hourly values of surface anthropogenic emissions are available for 15 primary pollutants:  $\text{NO}$ ,  $\text{NO}_2$ ,  $\text{CO}$ ,  $\text{SO}_2$ ,  $\text{CH}_4$ , and the 10 nonmethane VOCs: ethane, *n*-butane, ethene, propene, *o*-xylene, formaldehyde, acetaldehyde, methyl ethyl ketone, methanol and ethanol. Isoprene is considered as the only primary biogenic VOC. Its emission parameterization follows the model of Guenther et al. (1993).

## 2.2. The studied pollution event: ESQUIF IOP2

The period selected for this study is the second Intensive Observation Period (hereafter called IOP2) of the ESQUIF project (Menut et al., 2000; Vautard et al., 2003), extending from 7 August 1998 to 11 August 1998 (hereafter called 7A98 and 11A98, respectively). The IOP2 is a fully documented pollution event with both surface measurements and aircraft flights above the Paris area. Moreover, different situations were observed during this period: a pure local production on 7A98 (see Fig. 1) shifted to a mixing of local production with a long-range transport of ozone on 9A98, 10A98 and 11A98. IOP2 was extensively studied and more details on dynamics and chemistry can be found, e.g. in Honoré et al. (2000), Derognat et al. (2003), Menut (2003) and Beekmann and Derognat (2003). In particular, it was shown that CHIMERE accurately reproduces the whole event. Comparisons with ground measurements done by AIRPARIF, the air quality network over the Paris area, showed a good agreement with the simulated concentration fields. For two selected surface stations, comparisons are displayed in Fig. 2. Even if some discrepancies occurred, CHIMERE exhibits good scores over long periods and is now the forecast model used over Paris. Daily results and scores may be viewed at <http://euler.lmd.polytechnique.fr/pioneer/>.

## 3. Elaboration of the air traffic emission inventory

### 3.1. Available data

The air traffic inventory used for this work was first elaborated by the AIRPARIF network. This inventory was built in a research context and the chosen year was 1998 for the three major commercial airports in the Paris area, that are CdG, Orly and Le Bourget (Fig. 1).

The total masses of VOCs,  $\text{CO}$ ,  $\text{NO}_x$ ,  $\text{CO}_2$  and  $\text{SO}_2$  were distributed by AIRPARIF among the four LTO phases. The standard landing and take-off (LTO) cycle is defined by the International Civil Aviation Organization (ICAO). Since pollutant concentrations and noise emissions vary with engine speed (and therefore with altitude), one rotation of a plane is decomposed in four

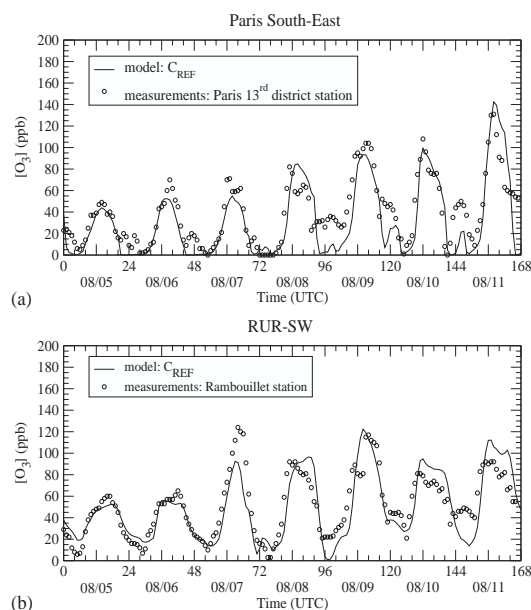


Fig. 2. Time series of ozone concentrations (ppb) simulated at [0–50] m and measured at surface in Paris southeast (13th district station) and in a rural area south-west of the domain (RUR-SW, Rambouillet station) from 5 to 11 August 1998.

phases (EPA, 1999): approach, idle or taxi, take-off and climbing. ICAO defined climbing as the interval between the end of take-off and the moment the plane exits the ABL and approach between the moment the plane enters the ABL and landing. ICAO's norms therefore take air traffic emissions into account from the surface to the top of the ABL, which height is defined at 915 m (3000 ft) by default. The altitude at which the motor power changes between take-off (maximum power) and climbing is also defined by default at 152 m (500 ft). For each phase, the emitted masses are spatially distributed among punctual sources, which cover the whole ground surface of the airports and represent aircraft trajectories, from the surface (5 m above ground) up to 915 m.

Time distribution of emissions at CdG and Orly was described by monthly, daily and hourly coefficients, which depend on the activity of each airport (and are considered as constant at Le Bourget).

Air traffic emission data were actually available as an inventory of punctual sources (reproducing trajectories) defined by total annual emitted masses (tons per year) of CO, CO<sub>2</sub>, NO<sub>x</sub>, total VOCs and SO<sub>2</sub>.

### 3.2. Adjusting annual emitted masses

Since the uncertainty on the provided estimated total masses was not available, they were compared to known

uncertainties of some others emission data. The only other available and comparable data set we found in the literature provides a quantification of SO<sub>2</sub>, VOCs and NO<sub>x</sub> emitted masses for the year of 1990 and a forecast for 2010 for North-American airports of various sizes (EPA, 1999). These masses were established by a “bottom-up” inventory, which takes into account the various plane and motor types and the number of completed rotations. This data set makes it possible to compare total masses emitted by Paris and American airports as a function of the number of LTOs, assuming differences between fleets are not too important. The results of this comparison are displayed in Fig. 3 for VOCs and NO<sub>x</sub>. The increase in emitted masses is not a linear function of the number of LTOs because queuing time (during which engines are in idle phase) increases rapidly with aircraft traffic. Best-fitting polynomials for

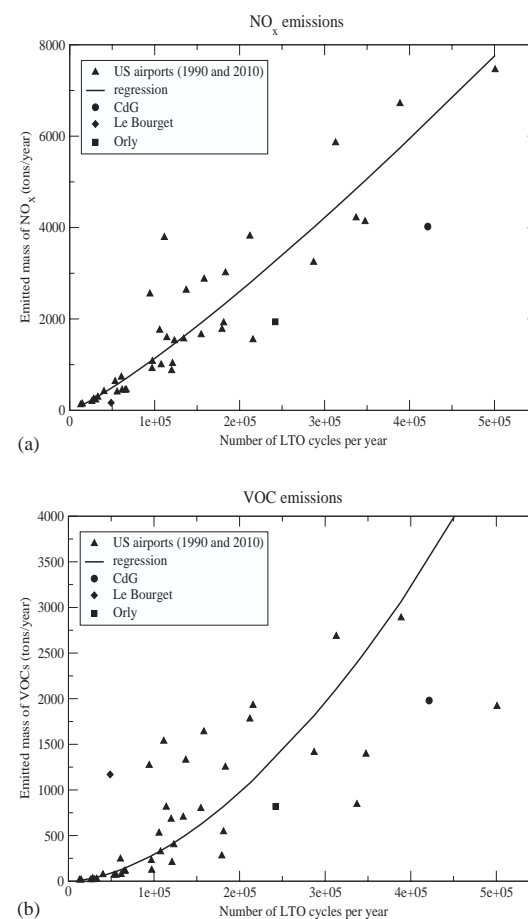


Fig. 3. NO<sub>x</sub> and VOC emissions as functions of the number of LTOs. Circle, square and diamond: 1998 data for the three airports of the Paris area.

Table 1

Multiplying coefficients applied to the emitted masses provided for Paris airports to be coherent with American airport emissions

	VOCs	NO <sub>x</sub>	SO <sub>2</sub>
CdG	1.78	1.57	0.51
Orly	1.65	1.68	0.51
Le Bourget	0.07	2.9	0.62

American data were thus computed. Since we can rely on the “bottom-up” method, multiplicative coefficients that bring the provided masses on the best-fitting curve for the same number of LTO cycles were deduced (detailed in Table 1) and applied to the provided emitted masses.

### 3.3. Spatial extension of the inventory

Since the top of the model reaches 3100 m, and since the boundary layer evolved from a few hundred meters to 2800 m during the studied case, the LTO standard altitude of 915 m has to be adapted for a satisfactory description of 3D emissions within our simulation domain. The inventory was thus extended from 915 up to 3100 m. We analyzed trajectories of commercial aircraft instrumented by the AMDAR project (AM-DAR, 2003) above the Paris area. This showed that a linear extension of trajectories is a satisfactory approximation. Since engine speed does not change during a whole phase, masses provided for climbing or approach were allocated to the extended trajectories.

### 3.4. Speciation

Through lack of information regarding aircraft, the speciation of NO<sub>x</sub> in order to retrieve NO and NO<sub>2</sub> masses was made using the usual “traffic” profile with 5% to NO<sub>2</sub> and 95% to NO in NO<sub>x</sub> mass. Because aircraft engine power differs during the landing and take-off, VOC speciation should not be constant during the whole LTO cycle. However, since no data on this point are available in the literature as confirmed by EEA (2001), it was impossible to specify different speciations for the various LTO cycle phases. VOC speciation was therefore based upon the only available average speciation provided by EEA (2001) for one commercial aircraft LTO cycle (Table 2). These species were distributed among the matching model species, retaining total VOC mass. Since the structure of the two siliceous compounds (13.3% of total mass) is unknown, they were assigned (according to their chemical formula only) to species that seem to reproduce at most their reactivity (i.e. alkanes/aromatics).

Table 2

Average speciation of VOC emissions for a commercial aircraft during one LTO cycle (EEA, 2001)

Real species	% of total VOC mass
Ethylene	19.2
Formaldehyde	16.6
C <sub>6</sub> H <sub>18</sub> O <sub>3</sub> Si <sub>3</sub>	10.1
Propene	5.7
Acetaldehyde	5.1
C <sub>8</sub> H <sub>24</sub> O <sub>4</sub> Si <sub>4</sub>	3.2
Ethyne	4.6
Acetone	2.6
Glyoxal	2.8
Acrolein	2.5
Butene	2.2
Benzene	2.1
1,3-Butadiene	2.0
Methyl glyoxal	2.2
<i>n</i> -Dodecane	1.2
Butyraldehyde	1.3
Others <1%	16.4
Others	0.2

### 3.5. Air traffic emissions compared to anthropogenic surface emissions

The contributions of aircraft emissions of our inventory are compared to anthropogenic surface emissions over the whole domain. Values are displayed in Table 3. The relative contribution of aircraft emissions to the total emitted mass in the domain is of +15.3% in NO<sub>x</sub> and +2.1% in VOCs, three of which are only emitted by aircraft (glyoxal, methyl glyoxal and acrolein).

Since there is a competition between VOCs and NO<sub>x</sub> for reacting with OH radicals, the VOC-to-NO<sub>x</sub> ratio of emissions is an indicator of their possible influence on ozone formation. When VOC-to-NO<sub>x</sub> ratio is high, radicals mainly react with VOCs and ozone production is finally sped up; when VOC-to-NO<sub>x</sub> ratio is low, radicals mainly react with NO<sub>2</sub> and ozone production is finally delayed (Seinfeld and Pandis, 1998).

In addition to their particular spatial distribution and in order to know whether air traffic emissions may have a different chemical influence on ozone concentrations, VOC-to-NO<sub>x</sub> ratios were computed. For air traffic emissions, they were estimated as averages for the three airports at each level for each hour during a weekday.

Thus, hourly VOC-to-NO<sub>x</sub> ratios at [0–50] m are displayed in Fig. 4(a). Since air traffic VOC-to-NO<sub>x</sub> ratio is an average, it varies with the relative activities of the three airports. Before 6 a.m. only Le Bourget airport has an important activity with aircraft that are very different from those of the other two airports.

Table 3

Emitted masses during a weekday (24 h) in the whole area (tons). The label “VOCs<sub>aircraft</sub>” correspond to the emissions of the three VOCs emitted only by aircraft, i.e. glyoxal, methylglyoxal and acrolein

Emissions (in tons/24 h)	CO	SO <sub>2</sub>	NO <sub>x</sub>	VOCs	
				Total	VOCs <sub>aircraft</sub>
Anthropogenic 2D	1798	117.5	275.5	776.6	
Aircraft 3D at [0–50] m	16.3	0.8	8.7	14.6	1
Aircraft 3D above 50 m	2.8	1.3	33.4	1.7	0.1

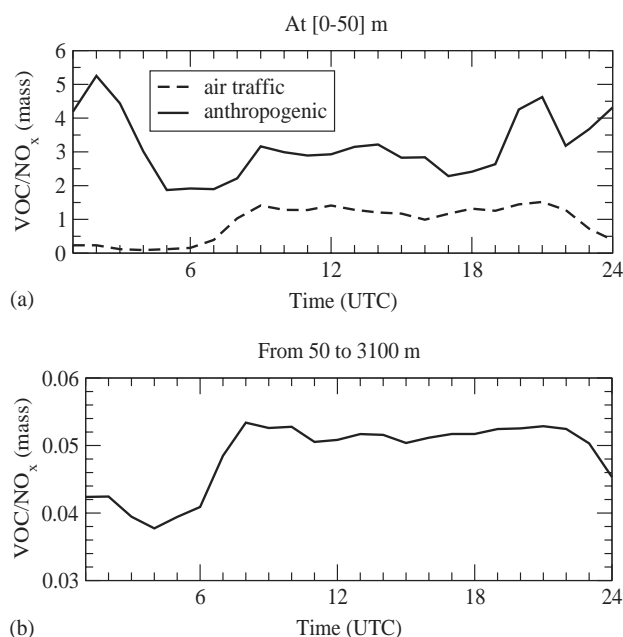


Fig. 4. VOC-to-NO<sub>x</sub> ratio for surface and air traffic emissions at [0–50] m (a) and average VOC-to-NO<sub>x</sub> ratios for air traffic emissions from 50 to 3100 m (b) during a weekday.

VOC-to-NO<sub>x</sub> ratios of air traffic emissions above 50 m are displayed in Fig. 4(b). At these altitudes, they are the only emissions taken into account in the modeled domain.

#### 4. Results

To estimate the impact of aircraft emissions on photo-oxidant pollution, two simulations of ESQUIF IOP2 were run:

- A first simulation hereafter called “the reference case” or  $C_{REF}$ , corresponding to the “classical” modeling of this situation (i.e. excluding aircraft emissions) as it was extensively explained by Vautard

et al. (2003) for the same period and strictly the same model version.

- A second simulation including air traffic emissions hereafter called  $C_{AT}$  and corresponding to the new parameterization built for this work.

In this paper, we compare  $C_{AT}$  to  $C_{REF}$  and discuss the influence of air traffic emissions in terms of impact on ozone concentrations, which is the classical indicator for photo-oxidant pollution.

In order to make the comparison between the two simulations easier, the contribution of air traffic emissions to ozone concentrations in the  $n$ th grid cell was calculated as the difference  $\Delta c_n$  between concentration values obtained for the case with air traffic emissions  $C_{AT}$  and the reference case  $C_{REF}$  ( $[c]_n^{AT}$  and  $[c]_n^{REF}$ ,



respectively):

$$\Delta c_n = [c]_n^{\text{AT}} - [c]_n^{\text{REF}}. \quad (1)$$

A positive difference indicates that ozone concentrations are higher when air traffic emissions are taken into account.

#### 4.1. Maximum impact

To estimate the maximum impact of air traffic emissions, horizontal maps of the extreme differences are displayed in Fig. 5. The results indicate, in each horizontal grid cell, the maximum and minimum differences in ozone concentrations during the whole simulated period and for all levels. Thus, these are not “instantaneous” maps but composite results. Negative differences are located close around the airports. This

decrease of ozone concentrations near the sources is due to the fast titration of ozone by NO. Positive differences are located in remote areas, mainly in the northwest part of the domain, far from the most intense sources such as Paris or the airports. This increase in ozone concentrations, which reaches up to 10 ppb, is due to the intakes of NO<sub>x</sub> in areas where the chemical system is NO-limited. The ozone net budget increases rapidly with moderate increases of NO (typically less than 1 ppb for the final concentration).

#### 4.2. Time distribution of maximum impact

To study the time occurrence of the maximum impacts of air traffic emissions on ozone concentrations, time series of  $\Delta c_n$  were examined at three locations, named rural area northeast (RUR-NE), rural area southwest (RUR-SW) and Paris southeast. These sites were chosen so that there is an urban area (Paris southeast) and two rural areas east and west of the domain. Moreover, previous studies showed that these sites, which correspond to AIRPARIF surface stations, are representative of their environment in terms of surface emitted fluxes and recorded ozone concentrations. Time series are displayed in Fig. 6 for three layers at [0–50], [50–650] and [650–1280] m.

In  $C_{\text{REF}}$ , at [0–50] m, an almost complete ozone titration is simulated during the night in Paris but not in RUR-NE and RUR-SW, which are representative of suburban environment with no significant local emissions of NO during the night (Fig. 6(a-1)). Peaks of concentration are observed in RUR-SW on 7A98 and 9A98, when this location is within the ozone plume flowing from Paris, and in RUR-NE on 10A98 and 11A98, when ozone is advected in the domain.

The strongest negative impact of air traffic emissions on ozone concentrations, which is due to ozone titration by NO, happened at night when photo-chemistry has no effect (Fig. 6(a-2)). At [0–50] m, for example, the decrease reaches –5.3 ppb in Paris southeast on 11A98 at 2 a.m. and –6.5 ppb in RUR-SW on 9A98 at 11 p.m.

The strongest positive impact of air traffic emissions on ozone concentrations happened during daylight, on 9A98, 10A98 and 11A98 (Fig. 6(a-2)). At [0–50] m, the increase reaches +3.8 ppb at 4 p.m. on 9A98 in RUR-SW and +6.5 ppb at 5 p.m. on 11A98 in RUR-NE.

On 7A98 only ozone concentrations are decreased in Paris during daylight (hours 55 to 68 in Fig. 6), when the town is downstream CdG, with a maximum at 3 p.m. of –8 ppb at [0–50] m and –9.5 ppb at [50–650] m. This decrease is associated with increases of +8 ppb over Paris and +6.5 ppb over RUR-SW at [650–1280] m at the same time (Fig. 6(c)). As indicated in Section 2.2, on 7A98 the pollution event was due to pure local production of ozone. This allows us to get a better view on the influence of the local air traffic emissions as the

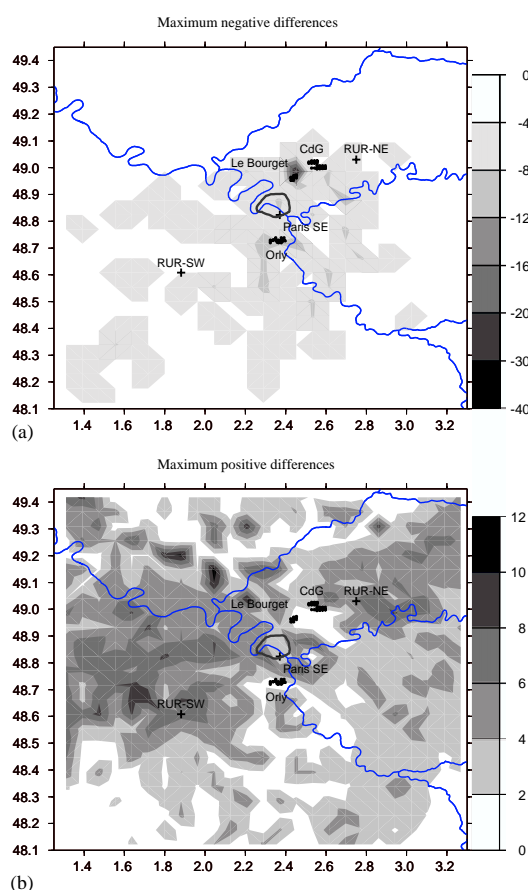


Fig. 5. Composite maps of differences in ozone concentrations between  $C_{\text{AT}}$  and  $C_{\text{REF}}$ . For each grid cell, the extreme difference (positive and negative) over the whole simulated period is represented in ppb.

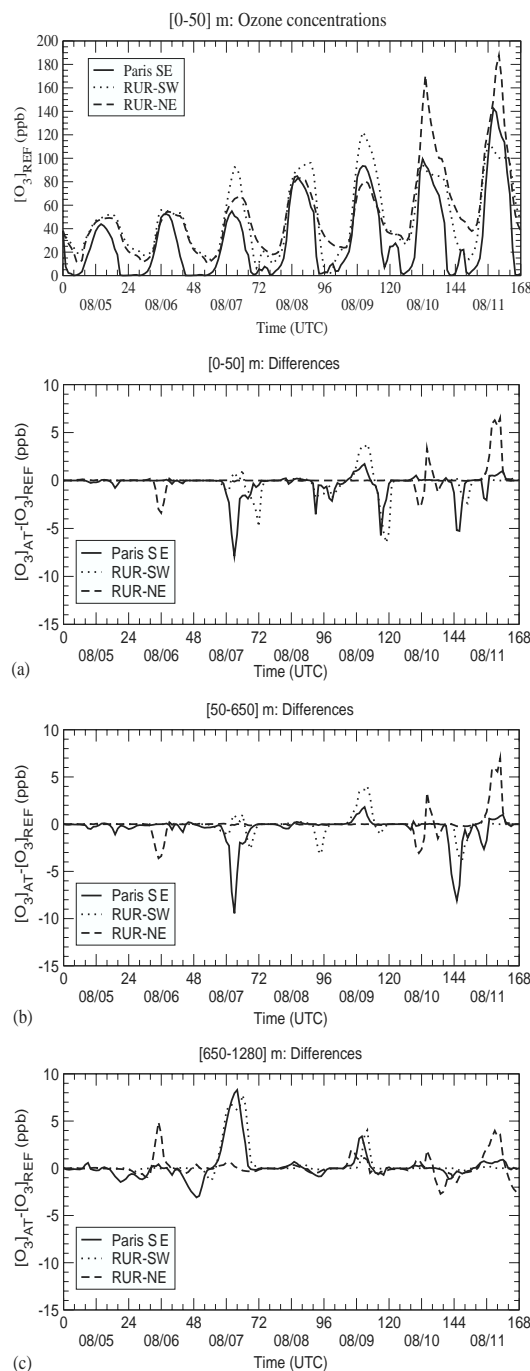


Fig. 6. Time evolution of  $\Delta c_n$  (difference in ozone concentrations between  $C_{AT}$  and  $C_{REF}$ ) over the whole simulated period in RUR-SW, Paris and RUR-NE (see Fig. 1 for location) at [0–50], [50–650] and [650–1280] m.

boundary conditions do not add noise. The case of 7A98 is therefore discussed hereafter in detail.

#### 4.3. Focus on 7A98

On this day a very particular pollution event took place: a well-marked ozone plume situation (Fig. 1 shows the situation at [0–50] m at 3 p.m.) was coupled to low wind speeds and a moderate maximum boundary layer height (BLH) (800 m at 3 p.m.). This episode is extensively described in Menut et al. (2000) and Menut (2003).

Since the wind direction was constant, the three particular locations are in the same situation during the whole day: a rural area upstream from the airport (RUR-NE), the urban area downstream two airports (Paris southeast) and a rural area downstream the three airports and the town (RUR-SW).

To get a better view of the vertical distribution of the impact of air traffic emissions during the daytime, the minimum and maximum differences at each level and at the three locations were retained (Fig. 7).

In opposition to the other days, the maximum impact of air traffic emissions near the surface (below 650 m) is a decrease in ozone concentrations: up to –3 ppb in rural areas and –9.5 ppb in urban areas. Since the BLH reaches 800 m at most on 7A98, the intense emissions due to aircraft moving at the ground are more concentrated than on other days.  $\text{NO}_x$  concentrations are then important enough to contribute to ozone titration downstream the airports. As an example, the decrease in ozone concentrations in Paris at [50–650] m corresponded to an increase in NO concentrations from 3 ppb in  $C_{REF}$  to 4.9 ppb in  $C_{AT}$ .

In altitude, the maximum impact is an increase in ozone concentrations downstream from the airports, in RUR-SW and in Paris where it reaches +8.5 ppb at 950 m. Since the BLH is very low, surface emissions are not mixed above 800 m and there are no other sources but aircraft. Thus, aircraft  $\text{NO}_x$  emissions take place in a layer that is not saturated in NO and contribute to ozone production, as previously seen in Section 4.1.

The 2.2 ppb decrease in ozone concentrations obtained in RUR-NE at 1650 m is due to the presence of a heavy traffic point where several plane trajectories cross in a small volume. Thus, at 8 p.m., when the decrease occurs, NO concentrations are not nil in  $C_{AT}$  as they were in  $C_{REF}$ . Since this is the end of the daytime, aircraft  $\text{NO}_x$  emissions are too important and contribute to ozone titration.

Since air traffic emissions decrease ozone concentrations near the surface, it is interesting to focus on what happens to the ozone peak that reaches 92 ppb in RUR-SW at 3 p.m. (see Fig. 1). To illustrate the situation better, Fig. 8 displays isoline of  $\Delta c_n$  at [0–50] m at 3 p.m. The plume is also marked by the



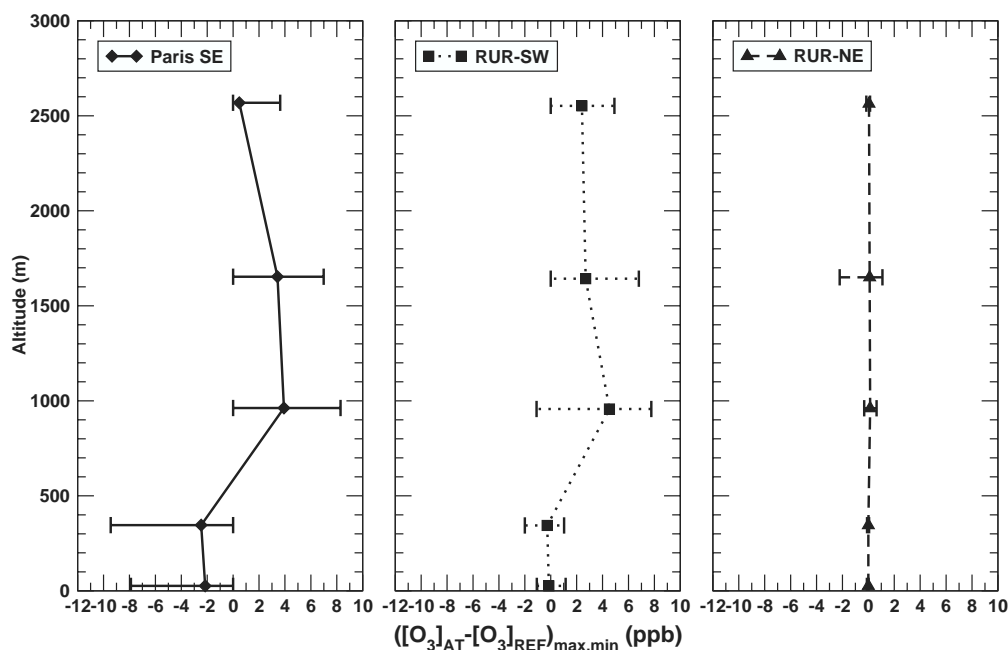


Fig. 7. Vertical profile of mean and extreme differences in ozone concentrations between  $C_{AT}$  and  $C_{REF}$  on 7A98 during daylight (from 7 a.m. to 8 p.m.).

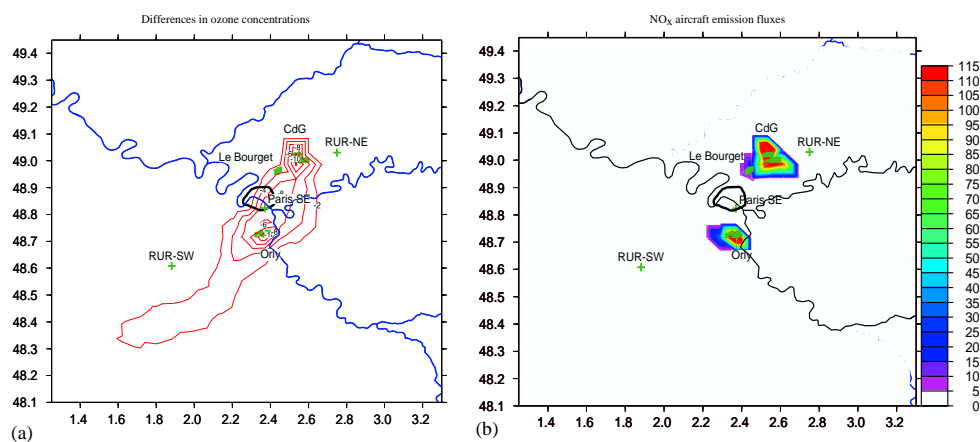


Fig. 8. Isolines of differences in ozone concentrations between  $C_{AT}$  and  $C_{REF}$  (a) and  $NO_x$  aircraft emission fluxes ( $\times 10^9$  molecules  $s^{-1} cm^{-2}$ ) (b) at [0–50] m on 7A98 at 3 p.m.

differences: ozone concentrations in the plume are decreased by  $-4$  to  $-2$  ppb. However, since concentrations are higher than 100 ppb in the core of the plume (Fig. 1), the variations due to air traffic emissions are too small to change the alert status of the event.

In the upper layers that were never mixed within the ABL during the day, photo-chemistry has acted on

aircraft emissions only to produce ozone during the afternoon. Since the atmosphere was very stable, changes in ozone concentrations are only due to local chemistry. As a consequence, isolines in Fig. 9 show the accumulation of ozone produced during the afternoon and the titration of ozone in areas where aircraft emit through this layer. At this time, after blowing from

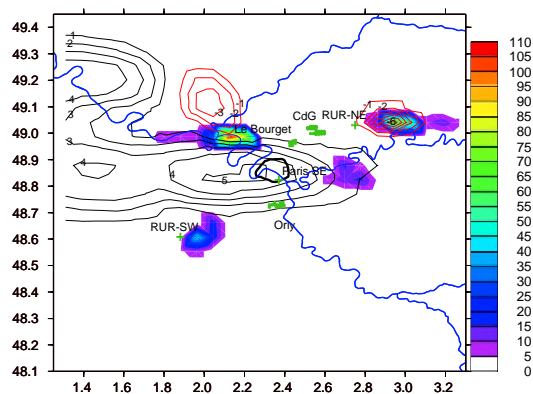


Fig. 9. Isolines of differences in ozone concentrations between  $C_{AT}$  and  $C_{REF}$  and NO aircraft emission fluxes ( $\times 10^9$  molecules  $s^{-1} cm^{-2}$ ) at [2030–3100] m on 7A98 at 10 p.m.

northeast the whole day, the wind just turned southeast and picked up west of Paris, which explains why the decrease due to the emissions is blown north.

#### 4.4. Summary of results

The maximum negative impact of air traffic emissions on ozone concentrations happens during the night and consists in a decrease of the already low ozone concentrations due to fast ozone titration by NO. Spatially, it is most important close to the sources, i.e. the airports, in the surface layer. The maximum positive impact of air traffic emissions on ozone concentrations happens during the daytime in remote areas and at altitude, where the chemical system is not saturated in NO.

On 7A98, when a pollution event took place, air traffic emissions have a negative impact on ozone concentrations in the ABL and a positive impact in the upper layers during the daytime. When the ABL remains very low, the intense emissions due to aircraft ground movements at the airports are more concentrated than on other days, which enhances ozone titration. In the upper layers, the atmosphere is relatively clean and does not naturally contain significant  $NO_x$  and VOC concentrations. Adding new emissions then enhanced photo-chemistry during the afternoon.

### 5. Relative impact of VOC and $NO_x$ aircraft emissions

As seen in the Section 3.2, the uncertainty on the specifically built emissions inventory was not available. Since this study is (to our knowledge) the first tentative performed over a large urban area, it is necessary to study whether variations of air traffic emissions, within a

reasonable range, could change the conclusions of the previous sections. In this last section, we thus examine the maximum variability of our results within the range uncertainties of the emissions, with the hypothesis that the uncertainty on air traffic emissions is of the same order of magnitude than the uncertainty estimated for surface emissions.

Since the relative impact of  $NO_x$  and VOCs has to be decoupled to be really understood (due to nonlinear chemical effects and chemical regimes occurring near urban areas), air traffic emitted masses of  $NO_x$  and VOCs were varied independently by more or less 30% and 40%, respectively, as it was done by Sillman et al. (2003) for surface emissions. Four new distinct simulations are therefore performed and compared to the previously described results. They have the same characteristics as  $C_{AT}$  with, respectively:

- air traffic VOC emissions reduced by 40%: case hereafter named  $C_{VOC-}$
- air traffic VOC emissions increased by 40%: case hereafter named  $C_{VOC+}$
- air traffic  $NO_x$  emissions reduced by 30%: case hereafter named  $C_{NO_x-}$
- air traffic  $NO_x$  emissions increased by 30%: case hereafter named  $C_{NO_x+}$

Differences in ozone concentrations were calculated for each case as in Section 4.

The effects of varying air traffic emissions are qualitatively predictable. In areas where the chemical system is saturated in NO, an increase of  $NO_x$  accentuates the decrease of ozone concentrations due to air traffic emissions, whereas a decrease of  $NO_x$  accentuates the increase of ozone concentrations. The evolution is opposite in nonsaturated areas. Since VOCs favor ozone production, ozone concentration decrease is accentuated when VOC emissions decrease and ozone concentration increase is accentuated when VOC emissions increase (Sillman et al., 2003).

We focus again on 7A98 to get a better view on what happens during the daytime, when ozone concentrations are high. The variations of minimum and maximum differences between  $C_{AT}$  and  $C_{REF}$  due to the perturbations of VOC or  $NO_x$  emissions were retained at each level at the three particular locations (Fig. 10(a) and (b), respectively).

Variations in VOC emissions only have an influence on the maximum impact downstream, in Paris southeast and Rur-SW near the surface (below 650 m).

The influence of variations in  $NO_x$  emissions is most important near the surface in Paris southeast, i.e. downstream CdG. Variations in  $NO_x$  emissions also have an influence on the maximum impact at altitude: downstream, the increase of 30% is small enough to keep the areas nonsaturated in NO; upstream, in Rur-NE, the decrease of 30% is small enough to keep

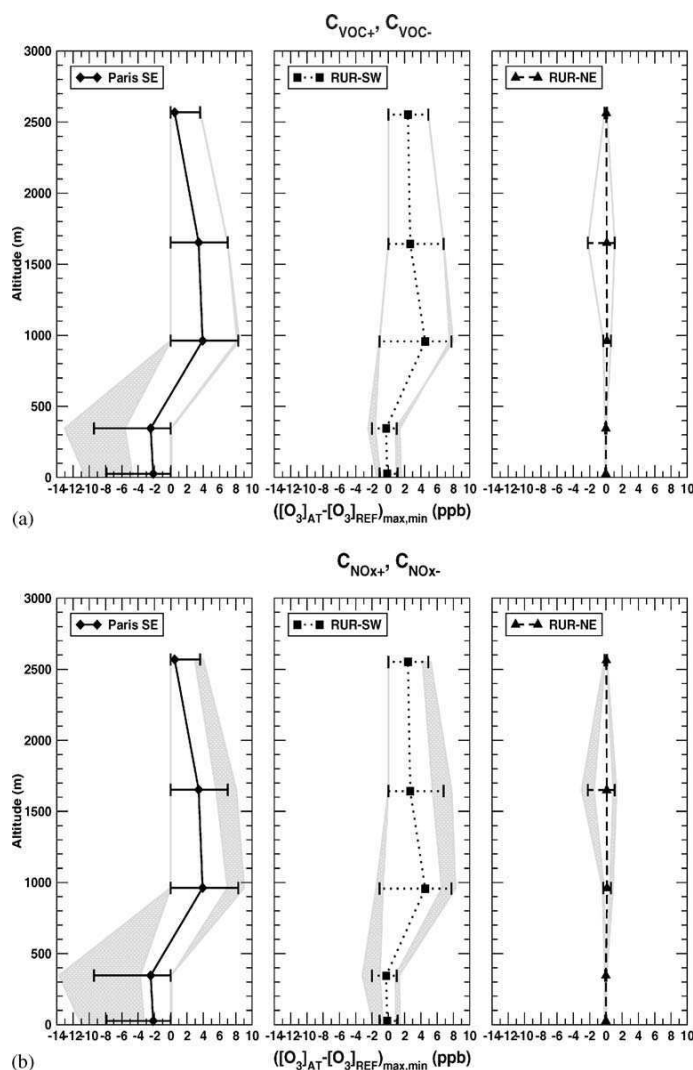


Fig. 10. Vertical profiles of mean and extreme differences in ozone concentrations between  $C_{AT}$  and  $C_{REF}$  with extreme variations of differences due to the perturbations of  $C_{NOx+}$ ,  $C_{NOx-}$ ,  $C_{VOC+}$ ,  $C_{VOC-}$  (gray shades) on 7A98 during daylight (from 7 a.m. to 8 p.m.).

the small volume where aircraft trajectories intersect saturated in NO.

Finally, variations of air traffic emissions within a range of uncertainty that is satisfactorily for surface emissions do not change the conclusions of the previous sections.

## 6. Conclusion

In this paper, we estimated the impact of air traffic emissions on simulated ozone concentrations. An air

traffic emission inventory was elaborated and simulations were carried out for the Paris area using CHIMERE CTM (reference  $C_{REF}$  without air traffic emissions and  $C_{AT}$  with air traffic emissions).

The maximum negative impact of air traffic emissions on ozone concentrations happens during the night and consists in a decrease of the already low ozone concentrations due to fast ozone titration by NO. Spatially, it is most important in the vicinity of the sources, i.e. the airports, in the surface layer. The maximum positive impact of air traffic emissions on ozone concentrations happens during the daytime in

remote areas and at altitude, where the chemical system is not saturated in NO.

On 7A98, during a pollution event, air traffic emissions have a negative impact on ozone concentrations in the ABL and a positive impact in the upper layers during the daytime. In the low ABL, the intense emissions due to aircraft ground movements at the airports are more concentrated than on other days. In the upper layers, the atmosphere is relatively clean and does not naturally contain significant NO<sub>x</sub> and VOC concentrations. Adding new emissions thus enhanced photo-chemistry during the afternoon. The produced ozone is vertically blocked within these stable layers and may be long-range transported.

The relative impact of VOC and NO<sub>x</sub> air traffic emissions was studied by varying independently VOC and NO<sub>x</sub> emitted masses. NO<sub>x</sub> air traffic emissions have the most important impact, particularly near the airports during the night, i.e. when ozone concentrations are very low. Increase of VOC air traffic emissions only have a significant impact during the 7A98 pollution event in the ABL. Nevertheless, variations of air traffic emissions within the same range of uncertainty as surface emissions do not change previous conclusions.

Our results suggest that aircraft emissions within the ABL have a nonnegligible impact on ozone concentrations in large urban areas. The net impact is negative, showing that airports induce an enhanced titration of ozone by NO. This impact is restricted to the close environment of the airports, certainly because of a fast chemistry combined to low wind speed (the most common case when a pollution event is observed in the Paris area).

### Acknowledgements

We thank Fanny MIETLICKI and Dominique GOMBERT (AIRPARIF: <http://www.airparif.asso.fr/>) for providing us with the three-dimensional aircraft emission inventory used in this paper.

### References

- AMDAR Project, 2003. <http://www.bom.gov.au/inside/oeb/networks/amdar.shtml>. Commonwealth Bureau of Meteorology.
- Beekmann, M., Derognat, C., 2003. Monte-Carlo uncertainty analysis of a regional scale transport chemistry model constrained by measurements from the ESQUIF campaign. *Journal of Geophysical Research—Atmospheres*, ESQUIF Special Section, 108, No. D17.
- Colville, R.N., Hutchinson, E.J., Mindell, J.S., Warren, F., 2001. The transport sector as a source of air pollution. *Atmospheric Environment* 35, 1537–1565.
- Derognat, C., Beekmann, M., Schmidt, H., Vautard, R., Martin, D., Neininger, B., 2003. Effect of biogenic VOC emissions on the tropospheric chemistry during elevated ozone periods in Ile-de-France. *Journal of Geophysical Research—Atmospheres*, ESQUIF Special Section, 108, No. D17.
- Derwent, R.G., 1982. Two-dimensional model studies of the impact of aircraft exhaust emissions on tropospheric ozone. *Atmospheric Environment* 16, 1997–2007.
- EPA (United States Environmental Protection Agency), 1999. Evaluation of Air Pollutant Emissions from Subsonic Commercial Jet Aircraft, EPA420-R-99-013. <http://www.epa.gov/otaq/aviation.htm>.
- European Environment Agency (EEA), 2001. EMEP/CORINAIR Emissions Inventory Guidebook, 3rd Edition. EEA, Group 08. <http://reports.eea.eu.int/EMEPCORINAIR3>.
- Gardner, R.M., Adams, K., Cook, T., Deidewig, F., Ernedal, S., Falk, R., Fleuti, E., Herms, E., Johnson, C.E., Lecht, M., Lee, D.S., Leech, M., Lister, D., Massé, B., Metcalfe, M., Newton, P., Schmitt, A., Vandenbergh, C., Van Drimmelen, R., 1997. The ANCAT/EC global inventory of NO<sub>x</sub> emissions from aircraft. *Atmospheric Environment* 31 (12), 1751–1766.
- Guenther, A.B., Zimmermann, P.R., Harley, P.C., Monson, R.K., Fall, R., 1993. Isoprene and monoterpene, model evaluation and sensitivity analysis. *Journal of Geophysical Research* 98, 12609–12617.
- Honoré, C., Vautard, R., Beekmann, M., 2000. Photochemical regimes in urban atmospheres: the influence of dispersion. *Geophysical Research Letters* 27 (13), 1895–1898.
- Lattuati, M., 1997. Contribution à l'étude du bilan de l'ozone troposphérique à l'interface de l'Europe et de l'Atlantique Nord: modélisation lagrangienne et mesures en altitude. Thèse de doctorat, Université Paris 6, France.
- Meijer, E.W., van Velthoven, P.F.J., Thompson, A.M., Pfister, L., Schlager, H., Schulte, P., Kelder, H., 2000. Model calculations of the impact of NO<sub>x</sub> from air traffic, lightning, and surface emissions, compared with measurements. *Journal of Geophysical Research* 105 (D3), 3833–3850.
- Menut, L., 2003. Adjoint modelling for atmospheric pollution processes sensitivity at regional scale during the ESQUIF IOP2. *Journal of Geophysical Research—Atmospheres*, ESQUIF Special Section, 108, No. D17.
- Menut, L., Vautard, R., Flamant, C., Abonne, C., Beekmann, M., Chazette, P., Flamant, P.H., Gombert, D., Guédalia, D., Lefebvre, M.P., Lossec, B., Martin, D., Mégie, G., Perros, P., Sicard, M., Toupance, G., 2000. Measurements and modelling of atmospheric pollution over the Paris area: an overview of the ESQUIF Project. *Annales Geophysicae* 18 (11), 1467–1481.
- Middleton, P., Stockwell, W.R., Carter, W.P., 1990. Aggregation and analysis of volatile organic compound emissions for regional modeling. *Atmospheric Environment* 24, 1107–1133.
- Moussiopoulos, N., Sahm, P., Karatzas, K., Papalexioiu, S., Karagiannidis, A., 1997. Assessing the impact of the new Athens airport to urban air quality with contemporary air pollution models. *Atmospheric Environment* 31 (10), 1497–1511.

- Perl, A., Patterson, J., Perez, M., 1997. Pricing aircraft emissions at Lyon-Satolas airport. *Transportation Research-D* 2 (2), 89–105.
- Sallès, J.J., Janischewski, A., Jaecker-Voirol, A., Martin, B., 1996. Mobile source emissions inventory model: application to the Paris area. *Atmospheric Environment* 30, 1965–1975.
- Schmidt, H., Martin, D., 2003. Continental scale adjoint sensitivity of ozone concentrations in the Paris region during the ESQUIF episodes. *Journal of Geophysical Research—Atmospheres*, ESQUIF Special Section, 108, No. D17.
- Schmidt, H., Derognat, C., Vautard, R., Beekmann, M., 2001. A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in western Europe. *Atmospheric Environment* 35 (36), 6277–6297.
- Seinfeld, J.H., Pandis, S.N., 1998. In: Relative roles of VOC and  $\text{NO}_x$  in ozone formation. *Atmospheric Chemistry and Physics*, from air pollution to climate change, 299–302. John Wiley & Sons Inc.
- Sillman, S., Vautard, R., Menut, L., Kley, D., 2003.  $\text{O}_3$ – $\text{NO}_x$ –VOC sensitivity and  $\text{NO}_x$ –VOC indicators in Paris: results from models and ESQUIF measurements. *Journal of Geophysical Research—Atmospheres*, ESQUIF Special Section, 108, No. D17.
- Vautard, R., Beekmann, M., Roux, J., Gombert, D., 2001. Validation of a hybrid forecasting system for the ozone concentrations over the Paris area. *Atmospheric Environment* 35 (14), 2449–2461.
- Vautard, R., Martin, D., Beekmann, M., Friedrich, R., Jaubertie, A., Kley, D., Lattuati, M., Moral, P., Neininger, B., Theloke, J., 2003. Paris emission inventory diagnostic from the ESQUIF airborne measurements and a chemistry-transport model. *Journal of Geophysical Research—Atmospheres*, ESQUIF Special Section, 108, No. D17.

# Références

- AIRPARIF (2004, Novembre). Évaluation de la qualité de l'air en Ile-de-France à l'échéance 2010 et impact du Plan de Protection de l'Atmosphère. Rapport technique. Étude réalisée pour le compte de la DRIRE Ile-de-France dans le cadre de l'élaboration du Plan de Protection de l'Atmosphère, commande DRIRE numéro 12 du 07/07/2003.
- Atkinson, R. (2000). Atmospheric chemistry of VOCs and  $\text{NO}_x$ . Atmospheric Environment **34**, 2063–2101.
- Baker, D. (2000). Inverse methods in global biogeochemical cycles, Chapitre An inversion method for determining time-dependent surface  $\text{CO}_2$  fluxes, pp. 279–294. AGU Press.
- Beekmann, M. et C. Derognat (2003). Monte-Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements from the Atmospheric Pollution Over the Paris Area (ESQUIF) campaign. Journal of Geophysical Research **108**(D17), 8559. doi :10.1029/2003JD003391.
- Bergamaschi, P., R. Hein, M. Heinmann, et P. Crutzen (2000). Inverse modeling of the global CO cycle : 1. Inversion of CO mixing ratio. Journal of Geophysical Research **105**(D2), 1909–1927.
- Bessagnet, B., A. Hodzic, R. Vautard, M. Beekmann, S. Cheinet, C. Honoré, C. Lioussé, et L. Rouil (2004). Aerosol modeling with CHIMERE - preliminary evaluation at the continental scale. Atmospheric Environment (38), 2803–2817.
- Blond, N. (2002a). Assimilation de données photochimiques et prévision de la pollution troposphérique. Physique et Chimie de l'atmosphère, Ecole Polytechnique, France. 202 p.
- Blond, N. (2002b). Assimilation de données photochimiques et prévision de la pollution troposphérique. Physique et Chimie de l'atmosphère, Ecole Polytechnique, France. 202 p.
- Blond, N., L. Bel, et R. Vautard (2003). Three-dimensional ozone data analysis with an air quality model over the Paris area. Journal of Geophysical Research **108**(D23), 4744.
- Blond, N. et R. Vautard (2004). Three-dimensional ozone data analyses and their use for short-term ozone prediction. Journal of Geophysical Research, in press.
- Bousquet, P., P. Ciais, P. Peylin, et P. Monfray (1999a). Inverse modeling and annual atmospheric  $\text{CO}_2$  sources and sinks : 1. Method and control inversion. Journal of Geophysical Research **104**(D21), 26161–26178.
- Bousquet, P., P. Ciais, P. Peylin, et P. Monfray (1999b). Inverse modeling and annual atmospheric  $\text{CO}_2$  sources and sinks : 2. Sensitivity study. Journal of Geophysical Research **104**(D21), 26179–26193.
- Brasseur, G., D. Hauglustaine, S. Walters, P. Rasch, J.-F. Müller, C. Granier, et X. Tie (1998). MOZART : A global three-dimensional chemical transport model for ozone and related chemical tracers. 1. Model description. Journal of Geophysical Research (103), 28265–28289.
- Chang, M., D. Hartley, C. Cardelino, D. Haas-Laursen, et W. Chang (1997). On using inverse methods for resolving emissions with large spatial inhomogeneities. Journal of Geophysical Research **102**(D13), 16023–16036.
- Chang, M., D. Hartley, et W.-L. Chang (1996). Inverse modeling of biogenic isoprene emissions. Geophysical Research Letters **23**(21), 3007.
- Cirpka, O. et P. Kitanidis (2001). Sensitivity of temporal moments calculated by the adjoint-state method and joint inverting of head and tracer data. Advances in Water Resources **24**, 89–103.
- Colella, P. et P. Woodward (1984). The piecewise parabolic method (PPM) for gas-dynamical simulations. Journal of Computational Physics (11), 38–39.
- Daley, R. (1996a). Atmospheric Data Analysis. Cambridge University Press.

- Daley, R. (1996b). Atmospheric Data Assimilation. Cambridge University Press.
- Daley, R. (1997). Atmospheric data assimilation. Journal of the Meteorological Society of Japan 75(1B), 319–329.
- Elbern, H. et H. Schmidt (1999). A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling. Journal of Geophysical Research 104, 18583–18598.
- Elbern, H., H. Schmidt, et A. Ebel (1997). Variational data assimilation for tropospheric chemistry modeling. Journal of Geophysical Research 102(D13), 15967–15985.
- Elbern, H., H. Schmidt, O. Talagrand, et A. Ebel (2000). 4D-variational data assimilation with an adjoint air quality model for emissions analysis. Environmental Modelling & Software 15, 539–548.
- Enting, I. (2002). Inverse problems in atmospheric constituent transport. Cambridge University Press.
- Enting, I. et J. Mansbridge (1991). Latitudinal distributions of sources and sinks of CO<sub>2</sub> : Results of an inversion study. Tellus 43B, 156–170.
- Erisman, J., W. van Pul, et G. Wyers (1994). Parametrization of surface resistance for the quantification of atmospheric deposition of acidifying pollutants and ozone. Atmospheric Environment (28), 2595–2607.
- ESQUIF (2001). ESQUIF Etude et Simulation de la Qualité de l'air en Ile-de-France, rapport final. Rapport technique, Institut Pierre-Simon Laplace, Laboratoire Interuniversitaire des Systèmes Atmosphériques, Météo-France, Laboratoire d'Aérodynamique, Forschungszentrum Jülich, AIRPARIF.
- EUROTRAC-2 (2001). GLOREAM, Global and Regional Atmospheric Modelling, Annual Report 2000. Rapport technique, International Scientific Secretariat GSF-Forschungszentrum für Umwelt und Gesundheit GmbH, München, Germany.
- Evensen, G. (1994, may). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. Journal of geophysical research 99(C5), 10143–10162.
- Finlayson-Pitts, B. et J. Pitts (2000). Chemistry of the upper and lower atmosphere. Academic Press.
- Fisher, M. et D. Lary (1996). Lagrangian four dimensional variational data assimilation of chemical species. Quarterly Journal of the Royal Meteorological Society 121, 1681.
- Fochesatto, J., P. Drobinski, C. Flamant, D. Guédalia, C. Sarrat, P. Flamant, et J. Pelon (2001). Evidence of dynamical coupling between the residual layer and the developing convective boundary layer. Boundary Layer Meteorology (99), 451–464.
- Gelb, A. (Ed.) (1974). Applied optimal estimation. Cambridge, Massachusetts : MIT Press.
- Gilbert, J.-C. et C. Lemaréchal (1989). Some numerical experiments with variable-storage quasi-Newton algorithms. Mathematical Programming 45, 407–435.
- Gilbert, J.-C. et C. Lemaréchal (1995, Juin). The modules M1QN3 and N1QN3, version 2.0c. Rapport technique, INRIA.
- Gilliland, A. B., R. L. Dennis, S. J. Roselle, et T. E. Pierce (2003). Seasonal NH<sub>3</sub> emission estimates for the eastern United States based on ammonium wet concentrations and an inverse modeling method. Journal of Geophysical Research 108(D15).
- Guenther, A. (1997). Seasonal and spatial variations in the natural volatile organic compound emissions. Ecological Applications (7), 34–45.
- Gurney, K., Y.-H. Chen, T. Maki, S. Kawa, A. Andrews, et Z. Zhu (2005). Sensitivity of atmospheric CO<sub>2</sub> inversions to seasonal and interannual variations in fossil fuel emissions. Journal of Geophysical Research 110.
- Haas-Laursen, D., D. Hartley, et R. Prinn (1996). Optimizing an inverse method to deduce time-varying emissions of trace gases. Journal of Geophysical Research 101(D17), 22823–22831.
- Hanna, S., J. Chang, et M. Fernau (1998). Monte-carlo estimates of uncertainties in predictions by a photochemical grid model (uam-iv) due to uncertainties in input variables. Atmospheric Environment 32, 3609–3618.
- Hanna, S., Z. Lu, H. Frey, N. Wheeler, J. Vukovitch, S. Arunachalam, M. Fernau, et D. Hansen (2001). Uncertainties in predicted ozone concentrations due to input uncertainties for the uam-v photochemical grid model applied to the July 1995 otag domain. Atmospheric Environment 35, 891–903.
- Hartley, D. et R. Prinn (1993). On the feasibility of determining surface emissions of trace gases using an inverse method in a three-dimensional chemical transport model. Journal of Geophysical Research 98, 5183–5198.

- Hassel, D., F. Weber, et T. Schmitz (2000). Zusammensetzung der Kohlenwasserstoffe im Abgas unterschiedlicher Fahrzeugkonzepte. Jülich, Germany.
- Heimann, M. et T. Kaminski (1999). Inverse modeling approaches to infer surface trace gas fluxes from observed atmospheric mixing ratios, Volume 14 of Approaches to scaling of trace gas fluxes in ecosystems. Elsevier, Amsterdam.
- Hein, R., P.J. Crutzen, et M. Heimann (1997). An inverse modeling approach to investigate the global atmospheric methane cycle. Global Biogeochemical Cycles (11), 43–76.
- Hicks, B., D. Baldocchi, T. Myers, R. Hosker, et D. Matt (1987). A preliminary multiple resistance routine for deriving dry deposition from measured quantities. Water, Air and Soil pollution (36), 311–330.
- Hodzic, A. (2002). Pollution atmosphérique en Ile-de-France. Master's thesis, Ecole Normale de la Météorologie.
- Hoelzemann, J., H. Elbern, et A. Ebel (2001). PSAS and 4D-var data assimilation for chemical state analysis by urban and rural observation sites. Physics and Chemistry of the Earth B 26(10), 807–812.
- Honoré, C. et R. Vautard (2000). Photochemical regimes in urban atmospheres : The influence of dispersion. Geophysical Research Letters 27, 1895–1898.
- Honoré, C., R. Vautard, et M. Beekmann (2000). Low and high  $\text{NO}_x$  chemical regimes in an urban environment. Environmental Modelling and Software, 559–564.
- Houweling, S., T. Kaminski, F. Dentener, J. Lelieveld, et M. Heimann (1999). Inverse modeling of methane sources and sinks using the adjoint of a global transport model. Journal of Geophysical Research 104(D21), 26137–26160.
- Jacob, D., B. Field, E. Jin, I. Bey, Q. Li, J. Logan, et R. Yantosca (2002). Atmospheric budget of acetone. Journal of Geophysical Research 107(D10), ACH5–1–19.
- Kalman, R. (1960). A new approach to linear filtering and prediction problems. Transactions of the ASME - Journal of Basic Engineering (82), 35–45.
- Kalman, R. et R. Bucy (1961). New results in linear filtering and prediction theory. Transactions of the ASME - Journal of Basic Engineering (83), 95–108.
- Kaminski, T. et M. Heimann (2001). Inverse modeling of atmospheric carbon dioxide fluxes. Science 294, 259.
- Kaminski, T., P. Rayner, M. Heimann, et I. Enting (2001). On aggregation errors in atmospheric transport inversions. Journal of Geophysical Research 106(D5), 4703.
- Kasibhatla, P., M. Heimann, P. Rayner, N. Mahowald, R. Prinn, et D. Hartley (2000). Inverse methods in global biogeochemical cycles. AGU Press.
- Kleinman, L. (1997). Dependence of ozone production on NO and hydrocarbons in the troposphere. Geophysical Research Letters 24(18), 2299–2302.
- Kühlwein, J. et R. Friedrich (2000). Uncertainties of modelling emissions from road transport. Atmospheric Environment 34, 4603–4610.
- Lacarra, J.-F. et O. Talagrand (1988). Short-range evolution of small perturbations in barotropic model. Tellus 40A, 81–95.
- Lanser, D. et J. Verwer (1998). Analysis of operator splitting for advection - diffusion - reaction problems. In Air Pollution modeling and Simulation Conference, Champs-sur-Marne, France.
- Lattuati, M. (1997). Contribution à l'étude du bilan de l'ozone troposphérique à l'interface de l'Europe et de l'Atlantique Nord : modélisation lagrangienne et mesures en altitude. Thèse de sciences, Université Paris 6, France.
- Louis, J. (1979). A parametric model of vertical eddy fluxes in the atmosphere. Boundary Layer Meteorology (17), 187–202.
- Louis, J., M. Tiedke, et J. Geleyn (1982). A short history of the PBL parametrization at ECMWF. In ECMWF Workshop on Planetary Boundary Layer parametrization, Reading, pp. 59–80.
- Lyster, P., S. Cohn, R. Menard, L.-P. Chang, S.-J. Lin, et R. Olsen (1997). An implementation of a two-dimensional filter for atmospheric chemical constituent assimilation of massively parallel computers. Monthly Weather Review 125, 1674.



- Madronich, S. et S. Flocke (1998). Handbook of Environment Chemistry, Chapitre The role of solar radiation in atmospheric chemistry, pp. 1–26. Heidelberg : P. Boule, Springer & Verlag.
- Mahowald, N., R. Prinn, et P. Rasch (1997). Deducing ccl3 f emissions using an inverse method and chemical transport models with assimilated winds. Journal of Geophysical Research **102**, 28153–28168.
- Ménard, R., S. Edouard, S. Houweling, G. Pétron, C. Granier, et C. Reeves (2004). Emissions of atmospheric trace compounds, Chapitre Data Assimilation and Inverse Methods. Kluwer Academic Publishers.
- Mendoza-Dominguez, A. et A. Russell (2001). Estimation of emission adjustments from the application of four-dimensional data assimilation to photochemical air quality modeling. Atmospheric Environment **35**, 2879–2894.
- Menut, L. (2003). Adjoint modelling for atmospheric pollution processes sensitivity at regional scale during the ESQUIF IOP2. Journal of Geophysical Research **108**(D17), 8562. doi :10.1029/2002JD002549.
- Menut, L., R. Vautard, M. Beekmann, et C. Honoré (2000). Sensitivity of photochemical pollution using the adjoint of a simplified chemistry-transport model. Journal of Geophysical Research **105**(D12), 15379–15402.
- Menut, L., R. Vautard, C. Flamant, C. Abonnel, M. Beekmann, P. Chazette, P. Flamant, D. Gombert, D. Guédalia, M. Lefebvre, B. Lossec, M. D., G. Mégie, P. Perros, M. Sicard, et G. Toupance (2000). Measurements and modelling of atmospheric pollution over the Paris area : an overview of the ESQUIF Project. Annales Geophysicae **18**(11), 1467–1481.
- Middleton, P., W. Stockwell, et W. Carter (1990). Aggregation and analysis of volatile organic compound emissions for regional modeling. Atmospheric Environment (24), 1107–1133.
- Mulholland, M. et J. Seinfeld (1995). Inverse air pollution modelling of urban-scale carbon monoxide emissions. Atmospheric Environment **29**, 497–516.
- Müller, J.-F. et T. Stavrakou (2005). Inversion of CO and NO<sub>x</sub> emissions using the adjoint of the IMAGES model. Atmospheric Chemistry and Physics **5**, 1157–1186.
- Ntziachristos, L., Z. Samaras, S. Eggleston, N. Gorissen, D. Hassel, A. Hickman, R. Joumard, R. R., et K. Zierock (2000). COPERT III Computer Programme to Calculate Emissions from Road Transport - Methodology and Emission Factors. Rapport technique, European Environment Agency.
- Pétron, G., C. Granier, B. Khatatov, J.-F. Lamarque, V. Yudin, J.-F. Müller, et J. Gille (2002). Inverse modeling of carbon monoxide surface emissions using Climate Monitoring and Diagnostics Laboratory network observations. Journal of Geophysical Research **107**(D24).
- Pétron, G., C. Granier, B. Khatatov, V. Yudin, J.-F. Lamarque, L. Emmons, J. Gille, et D. Edwards (2004). Monthly CO surface sources inventory based on the 2000-2001 MOPITT satellite data . Geophysical Research Letters **31**.
- Pison, I. et L. Menut (2004). Quantification of the impact of aircraft traffic emissions on tropospheric ozone over Paris area. Atmospheric Environment **38**(7), 971–983.
- Prinn, R. et D. Hartley (1995). Progress and Problems in Atmospheric Chemistry, Chapitre 6 - inverse methods in atmospheric chemistry, pp. 172–197. World Sciences.
- Quélo, D., V. Mallet, et B. Sportisse (2005). Inverse modeling of NO<sub>x</sub> emissions at regional scale over Northern France. Preliminary investigation of the second-order sensitivity. Journal of Geophysical Research.
- Reeves, C. and Cunnold, D., R. Derwent, E. Dlugodenczy, S. Edouard, C. Granier, R. Ménard, P. Novelli, et D. Parrish (2004). Emissions of atmospheric trace compounds, Chapitre Determination of Emissions from Observations of Atmospheric Compounds. Kluwer Academic Publishers.
- Rödenbeck, C., S. Houweling, M. Gloor, et M. Heimann (2003). CO<sub>2</sub> flux history 1982-2001 inferred from atmospheric data using a global inversion of atmospheric transport. Atmospheric Chemistry And Physics **3**, 1919–1964.
- Rudd, H. et I. Marlowe (1998). Improvements to the VOC inventory through speciation of white spirits : a report produced for the department of the environment, transport and the regions. Rapport technique, AEA Technology, Oxfordshire, England.

- Schmidt, H., C. Derognat, R. Vautard, et M. Beekmann (2001). A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in western Europe. *Atmospheric Environment* **35**, 6277–6297.
- Schmidt, H. et D. Martin (2003). Continental scale adjoint sensitivity of ozone concentrations in the Paris region during the ESQUIF episodes. *Journal of Geophysical Research* **108**(D17), 4313.
- Schmitz, T., D. Hassel, et F. Weber (2000). Determination of VOC components in the exhaust of gasoline and diesel passenger cars. *Atmospheric Environment* (34), 4639–4647.
- Sillman, S. (1999). The relation between ozone,  $\text{NO}_x$  and hydrocarbons in urban and polluted rural environments. *Atmospheric Environment* **33**, 1821–1845.
- Sillman, S., R. Vautard, L. Menut, et D. Kley (2003).  $\text{O}_3$ - $\text{NO}_x$ -VOC sensitivity indicators in Paris : results from models and ESQUIF measurements. *Journal of Geophysical Research* **108**(D17).
- Simpson, D. (1999). Inventorying emissions from nature in Europe. *Journal of Geophysical Research* **104**, 8113–8152.
- Society, E. I. (1994). Generation of European Emission Data for Episodes (GENEMIS) project, EUROTRAC annual report 1993, part 5. Rapport technique, EUROTRAC, Garmish-Partenkirchen, Germany.
- Stull, R. (1988). *An introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers.
- Talagrand, O. (1997). Assimilation of observations : an introduction. *Journal of the Meteorological Society of Japan* **75**(1B), 191–209.
- Theloke, J., A. Obermeier, et R. Friedrich (2000). Ermittlung des Lösemittelemissionen 1994 in Deutschland und Methoden zur Fortschreibung. In *Auftrag des Umweltbundesamtes*, Germany. Universität Stuttgart.
- Theloke, J., A. Obermeier, et R. Friedrich (2001). Abschätzung des Lösemittelemissionen in Deutschland. *Gefahrstoffe Reinhaltung Luft* (61), 105–111.
- Toupance, G., P. Perros, M. Soedomo, et N. Tsalkani (1988). Recherche sur la pollution oxydante à l'échelle régionale : le cas du bassin parisien. Rapport technique, Centre d'Étude et de Recherche sur la Pollution Atmosphérique Régionale, Université Paris-Val-de-Marne, Créteil, France.
- van de Velde, R., W. Faber, V. Katwijk, H. Scholten, T. Thewessen, M. Verspuy, et M. Zevenbergen (1994). The preparation of a European land-use data base. Rapport technique, RIVM Report 712401001, Bilthoven, The Netherlands.
- Van Leer, B. (1994). Upwind-difference methods for aerodynamic problems governed by the Euler equations. In *Lectures in Applied Mathematics*, Vol. 22, Part 2, Providence, RI, USA, pp. 327–336. American Mathematical Society.
- Vautard, R., M. Beekmann, I. Deleuze, et C. Honoré (1997). La pollution photochimique en région parisienne simulée par le modèle CHIMERE et l'influence du transport régional d'ozone. Rapport technique, Laboratoire de Météorologie Dynamique, 24 rue Lhomond 75231 Paris Cedex 05, France.
- Vautard, R., M. Beekmann, J. Roux, et D. Gombert (2001). Validation of a hybrid forecasting system for the ozone concentrations over the Paris area. *Atmospheric Environment* **35**, 2449–2461.
- Vautard, R., D. Martin, M. Beekmann, P. Drobinski, R. Friedrich, A. Jaubertie, D. Kley, M. Lattuat, P. Moral, B. Neininger, et J. Theloke (2003). Paris emission inventory diagnostics from ESQUIF airborne measurements and a chemistry transport model. *Journal of Geophysical Research* **108**(D17), 8564. doi :10.1029/2002JD002797.
- Vautard, R., L. Menut, M. Beekmann, P. Chazette, P. Flamant, D. Gombert, D. Guédalia, D. Kley, M. Lefèvre, D. Martin, G. Mégie, P. Perros, et G. Toupance (2003). A synthesis of the Air Pollution Over the Paris Region (ESQUIF) field campaign. *Journal of Geophysical Research* **108**(D17), 8558. doi :10.1029/2003JD003380.
- Vingarzan, R. (2004). A review of surface ozone background levels and trends. *Atmospheric Environment* **38**, 3431–3442.
- Wang, Y. et S. Bentley (2002). Development of a spatially explicit inventory of methane emissions from Australia and its verification using atmospheric concentration data. *Atmospheric Environment* **36**, 4965–4975.
- Wang, Y., M. McElroy, T. Wang, et P. Palmer (2004). Asian emissions of CO and  $\text{NO}_x$  : Constraints from aircraft and Chinese station data. *Journal of Geophysical Research* **109**.

- Wesely, M. (1989). Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. Atmospheric Environment (23), 1293–1304.
- Wesely, M. et B. Hicks (1977). Some factors that affect the deposition rates of sulfur dioxide and similar gases on vegetation. Journal of Air Pollution Control Association (27), 1110–1116.

## Modélisation inverse pour l'optimisation des sources primaires de pollution atmosphérique à l'échelle régionale

La pollution atmosphérique à l'échelle régionale est le bilan des interactions entre des processus très différents : les émissions, la chimie, le transport, le mélange et le dépôt des espèces gazeuses. La prévision de la qualité de l'air nécessite donc l'utilisation de modèles, qui prennent en compte les émissions par le biais de cadastres. Les concentrations de polluants simulées dépendent fortement des émissions utilisées. Or, les cadastres utilisés pour les représenter comportent de grandes incertitudes. Comme il est difficile actuellement d'affiner leurs méthodes de construction, il reste la possibilité d'ajouter de l'information aux cadastres existant. L'optimisation des émissions utilise l'information contenue dans les mesures pour obtenir le cadastre qui minimise la différence entre les concentrations simulées et mesurées.

Une méthode d'inversion des émissions anthropiques à l'échelle régionale, utilisant les mesures de routine et basée sur le modèle CHIMERE et son adjoint, a été élaborée et validée. Une technique de krigeage permet d'utiliser de façon optimale les informations disponibles dans l'espace des concentrations. L'enchaînement de cycles krigeage-optimisation améliore la qualité des résultats. Une technique d'agrégation spatiale dynamique est utilisée pour réduire la dimension du problème. Les émissions de  $\text{NO}_x$  du cadastre élaboré par AIRPARIF pour l'Île-de-France ont été inversées pendant les étés 1998 et 1999, les épisodes de la campagne ESQUIF étant étudiés en détail. L'optimisation corrige des écarts importants entre concentrations mesurées et simulées. Cependant, d'une façon générale, le niveau de fiabilité des résultats diminue avec la densité du réseau de mesure. Les résultats présentant le plus haut niveau de confiance concernent donc les flux d'émission les plus intenses d'Île-de-France. Les corrections apportées à la masse moyenne émise dans l'ensemble du domaine et aux profils temporels correspondant sont en accord avec l'estimation de l'incertitude sur le cadastre pour les  $\text{NO}_x$  (15%) obtenue lors de la campagne ESQUIF.

### Inverse modeling for the optimization of primary sources of atmospheric pollution at a regional scale

Atmospheric pollution at a regional scale is the result of various interacting processes : emissions, chemistry, transport, mixing and deposition of gaseous species. The forecast of air quality is then performed by models, in which the emissions are taken into account through inventories. The simulated pollutant concentrations depend highly on the emissions that are used. Now inventories that represent them have large uncertainties. Since it would be difficult today to improve their building methodologies, there remains the possibility of adding information to existing inventories. The optimization of emissions uses the information that is available in measurements to get the inventory that minimizes the difference between simulated and measured concentrations.

A method for the inversion of anthropogenic emissions at a regional scale, using network measurements and based on the CHIMERE model and its adjoint, was developed and validated. A kriging technique allows us to optimize the use of the information available in the concentration space. Repeated kriging-optimization cycles increase the quality of the results. A dynamical spatial aggregation technique makes it possible to further reduce the size of the problem.

The  $\text{NO}_x$  emissions from the inventory elaborated by AIRPARIF for the Paris area were inverted during the summers of 1998 and 1999, the events of the ESQUIF campaign being studied in detail. The optimization reduces large differences between simulated and measured concentrations. Generally, however, the confidence level of the results decreases with the density of the measurement network. Therefore, the results with the higher confidence level correspond to the most intense emission fluxes of the Paris area. On the whole domain, the corrections to the average emitted mass and to the matching time profiles are consistent with the estimate of 15% obtained during the ESQUIF campaign.